PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N

A2

(11) International Publication Number: WO 00/65025

(43) International Publication Date: 2 November 2000 (02.11.00)

(21) International Application Number: PCT/US00/10828

(22) International Filing Date: 21 April 2000 (21.04.00)

(30) Priority Data:
60/130,691 23 April 1999 (23.04.99) US

(71) Applicant *(for all designated States except US)*: MERCK & CO., INC. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065–0907 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: AUSTIN, Christopher, P. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065–0907 (US). BAILEY, Wendy, J. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065–0907 (US). MA, Lei [CN/US]; 126 East Lincoln Avenue, Rahway, NJ 07065–0907 (US).

(74) Common Representative: MERCK & CO., INC.; 126 East Lincoln Avenue, Rahway, NJ 07065–0907 (US).

(81) Designated States: CA, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: DNA MOLECULES ENCODING HUMAN ENDOTHELIN CONVERTING ENZYME 3

Met Glu Pro Pro Tyr Ser Leu Thr Ala His Tyr Asp Glu Phe Gln Glu
Val Lys Tyr Val Ser Arg Cys Gly Ala Gly Gly Ala Arg Gly Ala Ser
Leu Pro Pro Gly Phe Pro Leu Gly Ala Ala Arg Ser Ala Thr Gly Ala
Arg Ser Gly Leu Pro Arg Trp Asn Arg Arg Glu Val Cys Leu Leu Ser
Gly Leu Val Phe Ala Ala Gly Leu Cys Ala Ile Leu Ala Ala Met Leu
Ala Leu Lys Tyr Leu Gly Pro Val Ala Ala Gly Gly Gly Ala Cys Pro
Glu Gly Cys Pro Glu Arg Lys Ala Phe Ala Arg Ala Ala Arg Phe Leu
Ala Ala Asn Leu Asp Ala Ser Ile Asp Pro Cys Gln Asp Phe Tyr Ser
Phe Ala Cys Gly Gly Trp Leu Arg Arg His Ala Ile Pro Asp Asp Lys
Leu Thr Tyr Gly Thr Ile Ala Ala Ile Gly Glu Gln Asn Glu Glu Arg
Leu Arg Arg Leu Leu Ala Arg Pro Gly Gly Pro Gly Gly Ala Ala
Gln Arg Lys Val Arg Ala Phe Phe Arg Ser Cys Leu Asp Met Arg Glu
Ile Glu Arg Leu Gly Pro Arg Pro Met Leu Glu Val Ile Glu Asp Cys
Gly Gly Trp Asp Leu Gly Gly Ala Glu Glu Arg Pro Gly Val Ala Ala
Arg Trp Asp Leu Asn Arg Leu Leu Tyr Lys Ala Gln Gly Val Tyr Ser
Ala Ala Ala Leu Phe Ser Leu Thr Val Ser Leu Asp Asp Arg Asn Ser
Ser Arg Tyr Val Ile Arg Ile Asp Gln Asp Gly Leu Thr Leu Pro Glu
Arg Thr Leu Tyr Leu Ala Gln Asp Glu Asp Ser Glu Lys Ile Leu Ala
Ala Tyr Arg Val Phe Met Glu Arg Val Leu Ser Leu Leu Gly Ala Asp
Ala Val Glu Gln Lys Ala Gln Glu Ile Leu Gln Val Glu Gln Gln Leu
Ala Asn Ile Thr Val Ser Glu Tyr Asp Asp Leu Arg Arg Asp Val Ser
Ser Met Tyr Asn Lys Val Thr Leu Gly Gln Leu Gln Lys Ile Thr Pro
His Leu Arg Trp Lys Trp Leu Leu Asp Gln Ile Phe Gln Glu Asp Phe
Ser Glu Glu Glu Glu Val Val Leu Leu Ala Thr Asp Tyr Met Gln Gln
Val Ser Gln Leu Ile Arg Ser Thr Pro His Arg Val Leu His Asn Tyr
Leu Val Trp Arg Val Val Val Val Leu Ser Glu His Leu Ser Pro Pro
Phe Arg Glu Ala Leu His Glu Leu Ala Gln Glu Met Glu Gly Ser Asp
Lys Pro Gln Glu Leu Ala Arg Val Cys Leu Gly Gln Ala Asn Arg His
Phe Gly Met Ala Leu Gly Ala Leu Phe Val His Glu His Phe Ser Ala
Ala Ser Lys Ala Lys Val Gln Gln Leu Val Glu Asp Ile Lys Tyr Ile
Leu Gly Gln Arg Leu Glu Glu Leu Asp Trp Met Asp Ala Glu Thr Arg
Ala Ala Ala Arg Ala Lys Leu Gln Tyr Met Met Val Met Val Gly Tyr
Pro Asp Phe Leu Leu Lys Pro Asp Ala Val Asp Lys Glu Tyr Glu Phe
Glu Val His Glu Lys Thr Tyr Phe Lys Asn Ile Leu Asn Ser Ile Arg
Phe Ser Ile Gln Leu Ser Val Lys Lys Ile Arg Gln Glu Val Asp Lys
Ser Thr Trp Leu Leu Pro Pro Gln Ala Leu Asn Ala Tyr Tyr Leu Pro
Asn Lys Asn Gln Met Val Phe Pro Ala Gly Ile Leu Gln Pro Thr Leu
Tyr Asp Pro Asp Phe Pro Gln Ser Leu Asn Tyr Gly Gly Ile Gly Thr
Ile Ile Gly His Glu Leu Thr His Gly Tyr Asp Asp Trp Gly Gly Gln
Tyr Asp Arg Ser Gly Asn Leu Leu His Trp Trp Thr Glu Ala Ser Tyr
Ser Arg Phe Leu Arg Lys Ala Glu Cys Ile Val Arg Leu Tyr Asp Asn
Phe Thr Val Tyr Asn Gln Arg Val Asn Gly Lys His Thr Leu Gly Glu
Asn Ile Ala Asp Met Gly Gly Leu Lys Leu Ala Tyr His Ala Tyr Gln
Lys Trp Val Arg Glu His Gly Pro Glu His Pro Leu Pro Arg Leu Lys
Tyr Thr His Asp Gln Leu Phe Phe Ile Ala Phe Ala Gln Asn Trp Cys
Ile Lys Arg Arg Ser Gln Ser Ile Tyr Leu Gln Val Leu Thr Asp Lys
His Ala Pro Glu His Tyr Arg Val Leu Gly Ser Val Ser Gln Phe Glu
Glu Phe Gly Arg Ala Phe His Cys Pro Lys Asp Ser Pro Met Asn Pro
Ala His Lys Cys Ser Val Trp (SEQ ID NO:2)

(57) Abstract

The present invention relates to human DNA molecules encoding the endothelin converting enzyme–3(ECE–3)protein, recombinant vectors comprising DNA molecules encoding ECE–3, recombinant host cells which contain a recombinant vector encoding ECE–3, the ECE–3 protein encoded by the DNA molecule, and methods of identifying selective agonists and antagonists of ECE–3.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

TITLE OF THE INVENTION

DNA MOLECULES ENCODING HUMAN ENDOTHELIN CONVERTING ENZYME 3

CROSS-REFERENCE TO RELATED APPLICATIONS

Not Applicable

STATEMENT REGARDING FEDERALLY-SPONSORED R&D

Not Applicable

REFERENCE TO MICROFICHE APPENDIX

Not Applicable

FIELD OF THE INVENTION

The present invention relates to human DNA molecules encoding endothelin converting enzyme 3 (ECE-3), a membrane bound metalloprotease which proteolytically cleaves endothelin precursors to their active form, recombinant vectors comprising DNA molecules encoding ECE-3, recombinant host cells which contain a recombinant vector encoding ECE-3, the ECE-3 protein encoded by the DNA molecule, and methods of identifying selective modulators of ECE-3.

BACKGROUND OF THE INVENTION

Endothelins constitute a small family of 21 amino acid peptides that possess a various biological activities. These peptides are characterized by cysteine residues at positions 1, 3, 11 and 15. Endothelin-1 (ET-1) was the first to be disclosed and has been shown to be a potent endothelium-derived vasoconstrictor (Yanagisawa

et al., 1988, *Nature* 332: 411-415). The three known members of the endothelin family: ET-1, Endothelin-2 (ET-2) and Endothelin-3 (ET-3)] are produced in various tissues and are known to interact with two distinct G-protein coupled receptors, ETA and ETB (Elshourbagy et al., 1993, J. Biol. Chem. 268: 3873-3879), which are expressed in various cell types.

The endothelins are produced from large prepropolypeptides of approximately 200 amino acids. These precursors are initially cleaved into inactive 38-41 amino acid intermediates referred to as big ET-1, big ET-2 and big ET-3, respectively (Seidah, et al., 1993, *Ann. NY Acad. Sci.* 680: 135-146). The carboxy-terminal portion of the big ETs are then proteolytically cleaved at a conserved Trp21-Val/Ile22 juncture to produce the active endothelin peptide (i.e., ET-1, ET-2 or ET-3, respectively). The enzymes responsible for proteolytic cleavage of the big ETs to produce active endothelins are referred to as endothelin-converting enzymes (ECEs). To date, two ECEs have been disclosed, endothelin-converting enzyme 1 (ECE-1) and endothelin-converting enzyme 2 (ECE-2).

U.S. Patent No. 5,231,166, issued to Masaki et al on July 27, 1993, discloses and claims endothelin- 2.

U.S. Patent No. 5,294,569, issued to Masaki et al on March 15, 1994, discloses and claims DNA molecules which encode endothelin- 2.

U.S. Patent No. 5,548,061, issued to Masaki et al on August 20, 1996, discloses and claims endothelin- 3.

U.S. Patent No. 5,811,263 issued to Masaki et al on September 22, 1998, discloses and claims DNA molecules which encode endothelin- 3.

U.S. Patent No. 5,688,640, issued to Yanagisawa on November 18, 1997, discloses and claims methods of screening for modulators of ECE-1 utilizing bovine ECE-1.

Xu et al (1994, *Cell* 78: 473-485) disclose a cDNA encoding bovine ECE-1 and the concomitant amino acid sequence of bovine ECE-1.

U.S. Patent No. 5,736,376, issued to Yanagisawa on April 7, 1998, discloses and claims compositions which comprise bovine ECE-2.

Emoto and Yanagisawa (1995, J. Biol. Chem. 270 (25): 15262-15268) disclose a cDNA encoding bovine ECE-2 and the concomitant amino acid sequence of bovine ECE-2.

There is a body of work showing that specific antagonists of endothelin may play crucial roles in vascular disease models (e.g., see Ohlstein, et al., 1994, *Proc. Natl. Acad. Sci.* 91: 8052-8056; Clozel et al., 1993, *Nature* 365: 759-761; Giaid et al., 1993, *N. Engl. J. Med.* 328:1732-1740; Douglas et al., 1994, *Trends Pharmcol. Sci* 15: 313-316). Such modulators of endothelin levels may be useful to treat diseases such as hypertension, atherosclerosis and vascular restenosis, myocardial ischemia, cereberal vasospasm and subarachnoid hermorrhage, congestive heart failure, diabetes, endotoxic shock, migraine, possibly Raynaud's phenomenon (for a review, see Ohlstein et al., Functions Mediated by Peripheral Endothelin Receptors: in *Endothelin Receptors: From the Gene to the Heart* , @ Chapter 6, pp. 109-185; CRC Press, Boca Raton, FL, 1995) as well as pulmonary diseases such as asthma, pulmonary hypertension and adult respiratory distress syndrome (see also, Jorkasky et al., The Role of Endothelin in Human Disease: Implications and Potential Therapeutic Intervention: in *Endothelin Receptors: From the Gene to the Heart* , @ Chapter 8, pp. 215-271; CRC Press, Boca Raton, FL, 1995).

Despite the identification of ECEs as described above, it would be advantageous to identify additional ECEs, especially novel human ECEs, that are active in the cascade of events involving human endothelins and endothelin receptors. As pointed out by Xu et al (id.), bovine ECE-1 and bovine ECE-2 possess a strong preference for the big ET-1 substrate over the big ET-2 or big ET-3, showing that additional ECEs may exist and may have difference substrate specificity. The present invention addresses and meets these needs by disclosing an isolated nucleic acid fragment which expresses a form of human ECE-3, recombinant vectors which house this nucleic acid fragment, recombinant host cells which expresses human ECE-3 and/or a biologically active equivalent, and method of using DNA molecules encoding human ECE-3 and/or recombinant human ECE-3 protein to select modulators of ECE-3 and other ECE forms which in turn directly effect endothelin production and endothelin receptor function.

SUMMARY OF THE INVENTION

The present invention relates to an isolated or purified nucleic acid molecule (polynucleotide) which encodes a novel human endothelin converting enzyme, ECE-3. The nucleic acid molecules of the present invention are substantially free from other nucleic acids.

The present invention relates to an isolated nucleic acid molecule (polynucleotide) which encodes mRNA which expresses a novel novel human endothelin converting enzyme, ECE-3, this DNA molecule comprising the nucleotide sequence disclosed herein as SEQ ID NO:1.

The present invention also relates to biologically active fragments or mutants of SEQ ID NO:1 which encodes mRNA expressing a novel human endothelin converting enzyme, ECE-3. Any such biologically active fragment and/or mutant will encode either a protein or protein fragment which at least substantially mimics the pharmacological properties of a wild-type ECE-3 protein, including but not limited to the ECE-3 protein as set forth in SEQ ID NO:2. Any such polynucleotide includes but is not necessarily limited to nucleotide substitutions, deletions, additions, amino-terminal truncations and carboxy-terminal truncations such that these mutations encode mRNA which express a protein or protein fragment of diagnostic, therapeutic or prophylactic use and would be useful for screening for agonists and/or antagonists for ECE-3 function.

A preferred aspect of this portion of the present invention is disclosed in Figure 1A-B and Figure 4A-F, a human cDNA molecule (SEQ ID NO:1) encoding a novel ECE-3 protein.

The isolated nucleic acid molecules of the present invention may include a deoxyribonucleic acid molecule (DNA), such as genomic DNA and complementary DNA (cDNA), which may be single (coding or noncoding strand) or double stranded, as well as synthetic DNA, such as a synthesized, single stranded polynucleotide. The isolated nucleic acid molecule of the present invention may also include a ribonucleic acid molecule (RNA).

The present invention also relates to recombinant vectors and recombinant hosts, both prokaryotic and eukaryotic, which contain the substantially purified nucleic acid molecules disclosed throughout this specification.

The present invention also relates to subcellular membrane fractions of the recombinant host cells (both prokaryotic and eukaryotic as well as both stably and transiently transformed cells) which contain the proteins encoded by the nucleic acids of the present invention. These subcellular membrane fractions will comprise either wild-type or mutant forms of human ECE-3 proteins at levels substantially above endogenous levels and hence will be useful in various assays described throughout this specification. The preferred eukaryotic subcellular membrane locations for the

ECEs of the present invention include the cell membrane and the intracellelular Golgi membrane.

The present invention also relates to a substantially purified form of the human ECE-3 protein, which comprises the amino acid sequence disclosed in Figure 2A-B and Figure 4A-F and set forth as SEQ ID NO:2.

A preferred aspect of this portion of the present invention is human ECE-3, which consists of the amino acid sequence as set forth in SEQ ID NO:2 and Figure 2A-B and Figure 4A-F.

The present invention also relates to biologically active fragments and/or mutants of human ECE-3, comprising the amino acid sequence as set forth in SEQ ID NO:2, including but not necessarily limited to amino acid substitutions, deletions, additions, amino terminal truncations and carboxy-terminal truncations such that these mutations provide for proteins or protein fragments of diagnostic, therapeutic or prophylactic use and would be useful for screening for selective modulators, including but not limited to agonists and/or antagonists for ECE-3 function.

A preferred aspect of the present invention is disclosed in Figure 2A-B and Figure 4A-F and is set forth as SEQ ID NO:2, the amino acid sequence of the novel human ECE-3. Endothelin converting enzyme-3 is a novel member of ECE family of enzymes involved in the cleavage of big ET precursors to the active endothelin forms, ET-1, ET-2 and ET-3. The isolation, characterization and disclosure of a human form of a novel member of the ECE family will allow for more sophisticated methods of identifying selective modulators of the endothelin pathway in humans. The potential disease targets are exhaustive [e.g., see Xu et al (1994, *Cell* 78: 473-485) for a listing of pertinent literature], with the involvement of endothelins in systemic hypertension being an especially preferred area of concentration. Therefore, while other endothelin-converting enzymes have been isolated in the past, they have not accounted for all the production or actions of endothelin. The advantage of this invention is that it identifies a new member of the family of biosynthetic enzymes responsible for endothelin production, and thus represents a novel potential drug target. The products and method of the present invention will therefore also be useful for the study of endothelin production, metabolism, and biology as well as being useful in a potential compound screen for inhibitors of the

enzyme that may be therapeutic for various diseases mentioned herein, including but in no way limited to hypertension, vasospasm, or other vascular disorders.

The present invention also relates to polyclonal and monoclonal antibodies raised in response to either the human form of ECE-3, or a biologically active fragment thereof.

The present invention also relates to isolated nucleic acid molecules which are fusion constructions expressing fusion proteins useful in assays to identify compounds which modulate wild-type vertebrate ECE-3. A preferred aspect of this portion of the invention includes, but is not limited to, glutathione S-transferase (GST)-ECE-3 fusion constructs which include, but are not limited to, either the intracellular or intralumenal domain of human ECE-3 as an in-frame fusion at the carboxy terminus of the GST gene, or the extracellular and transmembrane ligand binding domain of ECE-3 fused to the amino terminus of GST, or the extracellular and transmembrane domain of ECE-3 fused to an immunoglobulin gene by methods known to one of ordinary skill in the art. Soluble recombinant GST-ECE-3 fusion proteins may be expressed in various expression systems, including *Spodoptera frugiperda* (Sf21) insect cells (Invitrogen) using a baculovirus expression vector (pAcG2T, Pharmingen). Such fusion constructs may also be useful in the generation of antibodies against ECE-3.

Therefore, the present invention relates to methods of expressing the human ECE-3 protein and biological equivalents disclosed herein, assays employing these gene products, recombinant host cells which comprise DNA constructs which express these proteins, and compounds identified through these assays which act as agonists or antagonists of ECE-3 activity or of another component of the endothelin pathway.

It is an object of the present invention to provide an isolated nucleic acid molecule (e.g., SEQ ID NO:1) which encodes a novel form of human ECE-3, or human ECE-3 fragments, mutants or derivatives of SEQ ID NO:2. Any such polynucleotide includes but is not necessarily limited to nucleotide substitutions, deletions, additions, amino-terminal truncations and carboxy-terminal truncations such that these mutations encode mRNA which express a protein or protein fragment of diagnostic, therapeutic or prophylactic use and would be useful for screening for selective modulators for vertebrate ECE-3 function.

It is a further object of the present invention to provide the human ECE-3 proteins or protein fragments encoded by the nucleic acid molecules referred to in the preceding paragraph.

It is a further object of the present invention to provide recombinant vectors and recombinant host cells which comprise a nucleic acid sequence encoding human ECE-3 or a biological equivalent thereof.

It is an object of the present invention to provide a substantially purified form of the human ECE-3 protein, as set forth in SEQ ID NO:2.

It is an object of the present invention to provide for biologically active fragments and/or mutants of the human ECE-3 protein, such as set forth in SEQ ID NO:2, including but not necessarily limited to amino acid substitutions, deletions, additions, amino terminal truncations and carboxy-terminal truncations such that these mutations provide for proteins or protein fragments of diagnostic, therapeutic and/or prophylactic use.

It is further an object of the present invention to provide for substantially purified subcellular fractions which comprise human ECE-3, especially subcellular fractions obtained from a host cell transfected or transformed with a DNA vector comprising a nucleotide sequence which encodes a protein which comprises the amino acid as set forth in SEQ ID NO:2 and Figure 2A-B and Figure 4A-F.

It is also an object of the present invention to provide for ECE-3-based in-frame fusion constructions, methods of expressing these fusion constructs, biological equivalents disclosed herein, related assays, recombinant cells expressing these constructs, and agonistic and/or antagonistic compounds identified through the use of the nucleic acid encoding vertebrate, mammalian and/or human ECE-3 protein as well as the expressed protein itself.

It is also an object of the present invention to use ECE-3 or membrane preparations containing ECE-3 or a biological equivalent to screen for modulators, preferably selective modulators, of ECE-3 activity. Any such compound may be useful in a diagnostic, therapeutic and/or prophylactic indications for such disease states as hypertension, atherosclerosis and vascular restenosis, myocardial ischemia, cereberal vasospasm, cerebral ischemia and subarachnoid hemorrhage, congestive heart failure, diabetes, benign prostatic hypertrophy, erectile dysfunction, renal disease and dysfunction, endotoxic shock, migraine, possibly Raynaud's phenomenon, as well

as pulmonary diseases such as asthma, pulmonary hypertension and adult respiratory distress syndrome.

As used herein, "substantially free from other nucleic acids" means at least 90%, preferably 95%, more preferably 99%, and even more preferably 99.9%, free of other nucleic acids. Thus, an ECE-3 DNA preparation that is substantially free from other nucleic acids will contain, as a percent of its total nucleic acid, no more than 10%, preferably no more than 5%, more preferably no more than 1%, and even more preferably no more than 0.1%, of non-ECE-3 nucleic acids. Whether a given ECE-3 DNA preparation is substantially free from other nucleic acids can be determined by such conventional techniques of assessing nucleic acid purity as, *e.g.*, agarose gel electrophoresis combined with appropriate staining methods, *e.g.*, ethidium bromide staining, or by sequencing.

As used herein, "substantially free from other proteins" or "substantially purified" means at least 90%, preferably 95%, more preferably 99%, and even more preferably 99.9%, free of other proteins. Thus, an ECE-3 protein preparation that is substantially free from other proteins will contain, as a percent of its total protein, no more than 10%, preferably no more than 5%, more preferably no more than 1%, and even more preferably no more than 0.1%, of non-ECE-3 proteins. Whether a given ECE-3 protein preparation is substantially free from other proteins can be determined by such conventional techniques of assessing protein purity as, *e.g.*, sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) combined with appropriate detection methods, *e.g.*, silver staining or immunoblotting. As used interchangeably with the terms "substantially free from other proteins" or "substantially purified", the terms "isolated ECE-3 protein" or "purified ECE-3 protein" also refer to ECE-3 protein that has been isolated from a natural source. Use of the term "isolated" or "purified" indicates that ECE-3 protein has been removed from its normal cellular environment. Thus, an isolated ECE-3 protein may be in a cell-free solution or placed in a different cellular environment from that in which it occurs naturally. The term isolated does not imply that an isolated ECE-3 protein is the only protein present, but instead means that an isolated ECE-3 protein is substantially free of other proteins and non-amino acid material (*e.g.*, nucleic acids, lipids, carbohydrates) naturally associated with the ECE-3 protein *in vivo*. Thus, an ECE-3 protein that is expressed in a prokaryotic or eukaryotic cell which do not naturally (*i.e.*, without human intervention) express it through recombinant means is

an "isolated ECE-3 protein." As noted above, an ECE-3 protein preparation that is an isolated or purified ECE-3 protein will be substantially free from other proteins will contain, as a percent of its total protein, no more than 10%, preferably no more than 5%, more preferably no more than 1%, and even more preferably no more than 0.1%, of non-ECE-3 proteins.

As used interchangeably herein, "functional equivalent" or "biologically active equivalent" means a protein which does not have exactly the same amino acid sequence as naturally occurring ECE-3, due to alternative splicing, deletions, mutations, substitutions, or additions, but retains substantially the same biological activity as ECE-3. Such functional equivalents will have significant amino acid sequence identity with naturally occurring ECE-3 and genes and cDNA encoding such functional equivalents can be detected by reduced stringency hybridization with a DNA sequence encoding naturally occurring ECE-3. For the purposes of this invention, naturally occurring ECE-3 has the amino acid sequence shown as SEQ ID NO:2 and is encoded by SEQ ID NO:1. A nucleic acid encoding a functional equivalent has at least about 50% identity at the nucleotide level to SEQ ID NO:1.

As used herein, "a conservative amino acid substitution" refers to the replacement of one amino acid residue by another, chemically similar, amino acid residue. Examples of such conservative substitutions are: substitution of one hydrophobic residue (isoleucine, leucine, valine, or methionine) for another; substitution of one polar residue for another polar residue of the same charge (e.g., arginine for lysine; glutamic acid for aspartic acid).

As used herein, "ECE" refers to -- endothelin converting enzyme --.

As used herein, "ECE-1" refers to -- endothelin converting enzyme-1 --.

As used herein, "ECE-2" refers to -- endothelin converting enzyme-2 --.

As used herein, "ECE-3" refers to -- endothelin converting enzyme-3 --.

As used herein, the term "mammalian host" will refer to any mammal, including a human being.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A-B show the nucleotide sequence which encodes human ECE-3, as set forth in SEQ ID NO:1.

Figure 2A-B show the amino acid sequence of human ECE-3, as set forth in SEQ ID NO:2.

Figure 3A-D show an autoradiograph of a Northern blot of tissue-specific human mRNA. This Northern analysis show the human ECE-3 gene to be expressed in medulla oblongata and ovary at a high level; putamen, spinal cord, caudate nucleus, substantia nigra, thalamus, and testis at a medium level; and amygdala, corpus callosum, hippocampus, whole brain, subthalamic nucleus, cerebellum, cerebral cortex, occipital pole, frontal lobe, temporal lobe, thymus, prostate, skeletal muscle, kidney, pancreas and heart at a low level.

Figure 4A-F show the coding (SEQ ID NO:1), and anticoding (SEQ ID NO:22) DNA sequence as well as the open reading frame and amino acid sequence (SEQ ID NO:2) human ECE-3.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to an isolated nucleic acid molecule (polynucleotide) which encodes a novel human endothelin converting enzyme, endothelin converting enzyme-3 (ECE-3). The nucleic acid molecules of the present invention are substantially free from other nucleic acids. For most cloning purposes, DNA is a preferred nucleic acid.

The present invention relates to an isolated nucleic acid molecule (polynucleotide) which encodes mRNA which expresses a novel human ECE-3, this DNA molecule comprising the nucleotide sequence disclosed herein as SEQ ID NO:1, shown herein as follows:

```
GGCGGCGGGC GCTGGGAGAC ACCGGACGCC CGCTCGGCTG CGCTGCGGCT CAGGCCCCCG
CTCGGGCCCG ACCCGCTCGG TCACCGCCGG CTCGGGCGCG CACCTGCCGG CTGCGGCCCC
AGGGCCATGC GGAGGCCCAC GAGGAGGCCG GCGGCCACGC GCATCCCGTA GCCCAGGTGG
CCCAGGTCTG CACCGCGGCG GCCTCGGCGC CATGGAGCCC CCGTATTCGC TGACGGCGCA
CTACGATGAG TTCCAAGAGG TCAAGTACGT GAGCCGCTGC GGCGCGGGGG GCGCGCGCGG
GGCCTCCCTG CCCCCGGGCT TCCCGTTGGG CGCTGCCCGC AGCGCCACCG GGGCCCGGTC
CGGGCTGCCG CGCTGGAACC GGCGCGAGGT GTGCCTGCTG TCGGGGCTGG TGTTCGCCGC
CGGCCTCTGC GCCATTCTGG CGGCTATGCT GGCCCTCAAG TACCTGGGCC CGGTCGCGGC
```

CGGCGGCGGC GCCTGTCCCG AGGGCTGCCC TGAGCGCAAG GCCTTCGCGC GCGCCGCTCG
CTTCCTGGCC GCCAACCTGG ACGCCAGCAT CGACCCATGC CAGGACTTCT ACTCGTTCGC
CTGCGGCGGT TGGCTGCGGC GCCACGCCAT CCCCGACGAC AAGCTCACCT ATGGCACCAT
CGCGGCCATC GGCGAGCAAA ACGAGGAGCG CCTACGGCGC CTGCTGGCGC GGCCCGGGGG
TGGGCCTGGC GGCGCGGCCC AGCGCAAGGT GCGCGCCTTC TTCCGCTCGT GCCTCGACAT
GCGCGAGATC GAGCGACTGG GCCCGCGACC CATGCTAGAG GTCATCGAGG ACTGCGGGGG
CTGGGACCTG GGCGGCGCGG AGGAGCGTCC GGGGGTCGCG GCGCGATGGG ACCTCAACCG
GCTGCTGTAC AAGGCGCAGG GCGTGTACAG CGCCGCCGCG CTCTTCTCGC TCACGGTCAG
CCTGGACGAC AGGAACTCCT CGCGCTACGT CATCCGCATT GACCAGGATG GGCTCACCCT
GCCAGAGAGG ACCCTGTACC TCGCTCAGGA TGAGGACAGT GAGAAGATCC TGGCAGCATA
CAGGGTGTTC ATGGAGCGAG TGCTCAGCCT CCTGGGTGCA GACGCTGTGG AACAGAAGGC
CCAAGAGATC CTGCAAGTGG AGCAGCAGCT GGCCAACATC ACTGTGTCAG AGTATGACGA
CCTACGGCGA GATGTCAGCT CCATGTACAA CAAGGTGACG CTGGGGCAGC TGCAGAAGAT
CACCCCCCAC TTGCGGTGGA AGTGGCTGCT AGACCAGATC TTCCAGGAGG ACTTCTCAGA
GGAAGAGGAG GTGGTGCTGC TGGCGACAGA CTACATGCAG CAGGTGTCGC AGCTCATCCG
CTCCACACCC CACCGGGTCC TGCACAACTA CCTGGTGTGG CGCGTGGTGG TGGTCCTGAG
TGAACACCTG TCCCCGCCAT TCCGTGAGGC ACTGCACGAG CTGGCACAGG AGATGGAGGG
CAGCGACAAG CCACAGGAGC TGGCCCGGGT CTGCTTGGGC CAGGCCAATC GCCACTTTGG
CATGGCGCTT GGCGCCCTCT TTGTACATGA GCACTTCTCA GCTGCCAGCA AAGCCAAGGT
GCAGCAGCTA GTGGAAGACA TCAAGTACAT CCTGGGCCAG CGCCTGGAGG AGCTGGACTG
GATGGACGCC GAGACCAGGG CTGCTGCTCG GGCCAAGCTC CAGTACATGA TGGTGATGGT
CGGCTACCCG GACTTCCTGC TGAAACCCGA TGCTGTGGAC AAGGAGTATG AGTTTGAGGT
CCATGAGAAG ACCTACTTCA AGAACATCTT GAACAGCATC CGCTTCAGCA TCCAGCTCTC
AGTTAAGAAG ATTCGGCAGG AGGTGGACAA GTCCACGTGG CTGCTCCCCC CACAGGCGCT
CAATGCCTAC TATCTACCCA ACAAGAACCA GATGGTGTTC CCCGCGGGCA TCCTGCAGCC
CACCCTGTAC GACCCTGACT TCCCACAGTC TCTCAACTAC GGGGGCATCG GCACCATCAT
TGGACATGAG CTGACCCACG GCTACGACGA CTGGGGGGGC CAGTATGACC GCTCAGGGAA
CCTGCTGCAC TGGTGGACGG AGGCCTCCTA CAGCCGCTTC CTGCGAAAGG CTGAGTGCAT
CGTCCGTCTC TATGACAACT TCACTGTCTA CAACCAGCGG GTGAACGGGA AACACACGCT
TGGGGAGAAC ATCGCAGATA TGGGCGGCCT CAAGCTGGCC TACCACGCCT ATCAGAAGTG
GGTGCGGGAG CACGGCCCAG AGCACCCACT TCCCCGGCTC AAGTACACAC ATGACCAGCT
CTTCTTCATT GCCTTTGCCC AGAACTGGTG CATCAAGCGG CGGTCGCAGT CCATCTACCT
GCAGGTGCTG ACTGACAAGC ATGCCCCTGA GCACTACAGG GTGCTGGGCA GTGTGTCCCA
GTTTGAGGAG TTTGGCCGGG CTTTCCACTG TCCCAAGGAC TCACCCATGA ACCCTGCCCA

CAAGTGTTCC GTGTGGTGAG CCTGGCTGCC CGCCTGCACG CCCCCACTGC CCCCGCACGA
ATCACCTCCT GCTGGCTACC GGGGCAGGCA TGCACCCGGT GCCAGCCCCG CTCTGGGCAC
CACCTGCCTT CCAGCCCCTC CAGGACCCGG TCCCCCTGCT GCCCCTCACT TCAGGAGGGG
CCTGGAGCAG GGTGAGGCTG GACTTTGGGG GGCTGTGAGG GAAATATACT GGGGTCCCCA
GATTCTGCTC TAAGGGGGCC AGACCCTCTG CCAGGCTGGA TTGTACGGGC CCCACCTTCG
CTGTGTTCTT GCTGCAAAGT CTGGTCAATA AATCACTGCA CTGTTAAAAA AAAAAAAAAA
AAAAAATTCC TGCG (SEQ ID NO:1).

The above-exemplified isolated DNA molecule, shown in Figure 1A-B and Figure 4A-F and set forth as SEQ ID NO:1, contains 2894 nucleotides. This DNA molecule contains an open reading frame from nucleotide 212 (initiating Met from nt 212-214) to nucleotide 2536, with a "TGA" termination codon from nucleotides 2537-2539. A Kozak sequence (GGCGCCATGG [contained from nt 206-215 of SEQ ID NO:1]) is present and a polyA+ site is evident from nt 2866-2886 of SEQ ID NO:1. This open reading frame encodes a preferred form of the present invention, a human ECE-3 protein. The ECE-3 protein contains an open reading frame of 775 amino acids in length, as shown in Figure 2A-B and Figure 4A-F and as set forth in SEQ ID NO:2. Radiation hybrid mapping assigns this gene to chromosome 2q37, in a region that has not been linked to any human disease that might logically be due to ECE mutation. Northern analysis show the gene to be expressed in medulla and ovary at a high level; putamen, spinal cord, testis, caudate nucleus, substantia nigra and thalamus at a medium level; and amygdala, corpus callosum, hippocampus, whole brain, subthalamic nucleus, cerebellum, cerebral cortex, occipital pole, frontal lobe, temporal lobe, thymus, prostate, skeletal muscle, kidney, pancreas and heart at a low level (Figure 3A-3D).

The full length cDNA encoding ECE-3 was isolated by initially searching a public EST database, wherein an EST corresponding to a third member of the endothelin family was found, and a full-length sequence was isolated by EST sequencing and RACE (Rapid Amplification of cDNA Ends). A unigene search looking for endothelin converting enzyme hits index class Hs.26880 which consists of 3 ESTs and is annotated as being 44% similar to the rat ECE-1. The three ESTs represent 2 Image clones (AA523527, [R61440, R61395]). R61440 is a 5' EST 606 bp long (SEQ ID NO: 3) which hits KIAA0604 with 60% similarity over only a portion of the full length EST (167/276 bp). No similar ESTs were found during this search of the public database. The two 3' EST sequences (SEQ ID NOs: 4 and 5) are

identical to each other and to no other EST. The nucleotide sequence of each EST (SEQ ID NOs: 3-5) are disclosed in Example Section 1 of this specification as well as from the National Center for Biotechnology Information (NCBI) homepage at http://www.ncbi.nlm.nih.gov/.

The present invention also relates to biologically active fragments or mutants of SEQ ID NO:1 which encodes mRNA expressing ECE-3. Any such biologically active fragment and/or mutant will encode either a protein or protein fragment which at least substantially mimics the enzymatic properties of human ECE-3 protein, including but not limited to the human ECE-3 protein as set forth in SEQ ID NO:2. Any such polynucleotide includes but is not necessarily limited to nucleotide substitutions, deletions, additions, amino-terminal truncations and carboxy-terminal truncations such that these mutations encode mRNA which express a protein or protein fragment of diagnostic, therapeutic or prophylactic use and would be useful for screening for agonists and/or antagonists for ECE-3 function and/or of modulators of other components of the endothelin receptor pathway.

A preferred aspect of this portion of the present invention is disclosed in Figure 1A-B, a cDNA molecule encoding human ECE-3 (SEQ ID NO:1).

The isolated nucleic acid molecules of the present invention may include a deoxyribonucleic acid molecule (DNA), such as genomic DNA and complementary DNA (cDNA), which may be single (coding or noncoding strand) or double stranded, as well as synthetic DNA, such as a synthesized, single stranded polynucleotide. The isolated nucleic acid molecule of the present invention may also include a ribonucleic acid molecule (RNA).

The degeneracy of the genetic code is such that, for all but two amino acids, more than a single codon encodes a particular amino acid. This allows for the construction of synthetic DNA that encodes the ECE-3 protein where the nucleotide sequence of the synthetic DNA differs significantly from the nucleotide sequence of SEQ ID NO:1, but still encodes the same ECE-3 protein as SEQ ID NO:1. Such synthetic DNAs are intended to be within the scope of the present invention. If it is desired to express such synthetic DNAs in a particular host cell or organism, the codon usage of such synthetic DNAs can be adjusted to reflect the codon usage of that particular host, thus leading to higher levels of expression of ECE-3 protein in the host. In other words, this redundancy in the various codons which code for specific amino acids is within the scope of the

present invention. Therefore, this invention is also directed to those DNA sequences which encode RNA comprising alternative codons which code for the eventual translation of the identical amino acid, as shown below:

A=Ala=Alanine: codons GCA, GCC, GCG, GCU
C=Cys=Cysteine: codons UGC, UGU
D=Asp=Aspartic acid: codons GAC, GAU
E=Glu=Glutamic acid: codons GAA, GAG
F=Phe=Phenylalanine: codons UUC, UUU
G=Gly=Glycine: codons GGA, GGC, GGG, GGU
H=His =Histidine: codons CAC, CAU
I=Ile =Isoleucine: codons AUA, AUC, AUU
K=Lys=Lysine: codons AAA, AAG
L=Leu=Leucine: codons UUA, UUG, CUA, CUC, CUG, CUU
M=Met=Methionine: codon AUG
N=Asp=Asparagine: codons AAC, AAU
P=Pro=Proline: codons CCA, CCC, CCG, CCU
Q=Gln=Glutamine: codons CAA, CAG
R=Arg=Arginine: codons AGA, AGG, CGA, CGC, CGG, CGU
S=Ser=Serine: codons AGC, AGU, UCA, UCC, UCG, UCU
T=Thr=Threonine: codons ACA, ACC, ACG, ACU
V=Val=Valine: codons GUA, GUC, GUG, GUU
W=Trp=Tryptophan: codon UGG
Y=Tyr=Tyrosine: codons UAC, UAU

Therefore, the present invention discloses codon redundancy which may result in differing DNA molecules expressing an identical protein. For purposes of this specification, a sequence bearing one or more replaced codons will be defined as a degenerate variation. Also included within the scope of this invention are mutations either in the DNA sequence or the translated protein which do not substantially alter the ultimate physical properties of the expressed protein. For example, substitution of valine for leucine, arginine for lysine, or asparagine for glutamine may not cause a change in functionality of the polypeptide.

It is known that DNA sequences coding for a peptide may be altered so as to code for a peptide having properties that are different than those of the naturally occurring peptide. Methods of altering the DNA sequences include

but are not limited to site directed mutagenesis. Examples of altered properties include but are not limited to changes in the affinity of an enzyme for a substrate or a receptor for a ligand.

The present invention also relates to recombinant vectors and recombinant hosts, both prokaryotic and eukaryotic, which contain the substantially purified nucleic acid molecules disclosed throughout this specification. The nucleic acid molecules of the present invention encoding ECE-3, in whole or in part, can be linked with other DNA molecules, i.e, DNA molecules to which the human ECE-3 are not naturally linked, to form "recombinant DNA molecules" which encode ECE-3. The novel DNA sequences of the present invention can be inserted into vectors which comprise nucleic acids encoding human ECE-3 or a functional equivalent. These vectors may be comprised of DNA or RNA; for most cloning purposes DNA vectors are preferred. Typical vectors include plasmids, modified viruses, bacteriophage, cosmids, yeast artificial chromosomes, and other forms of episomal or integrated DNA that can encode a ECE-3 protein. It is well within the skilled artisan to determine an appropriate vector for a particular gene transfer or other use.

Included in the present invention are DNA sequences that hybridize to SEQ ID NO:1 under stringent conditions. By way of example, and not limitation, a procedure using conditions of high stringency is as follows: Prehybridization of filters containing DNA is carried out for 2 hours to overnight at 65°C in buffer composed of 6X SSC, 5X Denhardt's solution, and 100 μg/ml denatured salmon sperm DNA. Filters are hybridized for 12 to 48 hrs at 65°C in prehybridization mixture containing 100 μg/ml denatured salmon sperm DNA and 5-20 X 10^6 cpm of ^{32}P-labeled probe. Washing of filters is done at 37°C for 1 hr in a solution containing 2X SSC, 0.1% SDS. This is followed by a wash in 0.1X SSC, 0.1% SDS at 50°C for 45 min. before autoradiography. Other procedures using conditions of high stringency would include either a hybridization step carried out in 5XSSC, 5X Denhardt's solution, 50% formamide at 42°C for 12 to 48 hours or a washing step carried out in 0.2X SSPE, 0.2% SDS at 65°C for 30 to 60 minutes.

Reagents mentioned in the foregoing procedures for carrying out high stringency hybridization are well known in the art. Details of the composition of these reagents can be found in, e.g., Sambrook et al., 1989, *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

In addition to the foregoing, other conditions of high stringency which may be used are well known in the art.

The present invention also relates to a substantially purified form of the human ECE-3 protein, which comprises the amino acid sequence disclosed in Figure 2A-B and as set forth in SEQ ID NO:2.

The present invention also relates to biologically active fragments and/or mutants of the human ECE-3 protein comprising the amino acid sequence as set forth in SEQ ID NO:2, including but not necessarily limited to amino acid substitutions, deletions, additions, amino terminal truncations and carboxy-terminal truncations such that these mutations provide for proteins or protein fragments of diagnostic, therapeutic or prophylactic use and would be useful for screening for agonists and/or antagonists of ECE-3 function or other components of the endothelin signal pathway.

A preferred aspect of the present invention is disclosed in Figure 2A-B and is set forth as SEQ ID NO:2 in three letter code, and as herein set forth as follows:

Met Glu Pro Pro Tyr Ser Leu Thr Ala His Tyr Asp Glu Phe Gln Glu
Val Lys Tyr Val Ser Arg Cys Gly Ala Gly Gly Ala Arg Gly Ala Ser
Leu Pro Pro Gly Phe Pro Leu Gly Ala Ala Arg Ser Ala Thr Gly Ala
Arg Ser Gly Leu Pro Arg Trp Asn Arg Arg Glu Val Cys Leu Leu Ser
Gly Leu Val Phe Ala Ala Gly Leu Cys Ala Ile Leu Ala Ala Met Leu
Ala Leu Lys Tyr Leu Gly Pro Val Ala Ala Gly Gly Gly Ala Cys Pro
Glu Gly Cys Pro Glu Arg Lys Ala Phe Ala Arg Ala Ala Arg Phe Leu
Ala Ala Asn Leu Asp Ala Ser Ile Asp Pro Cys Gln Asp Phe Tyr Ser
Phe Ala Cys Gly Gly Trp Leu Arg Arg His Ala Ile Pro Asp Asp Lys
Leu Thr Tyr Gly Thr Ile Ala Ala Ile Gly Glu Gln Asn Glu Glu Arg
Leu Arg Arg Leu Leu Ala Arg Pro Gly Gly Gly Pro Gly Gly Ala Ala
Gln Arg Lys Val Arg Ala Phe Phe Arg Ser Cys Leu Asp Met Arg Glu
Ile Glu Arg Leu Gly Pro Arg Pro Met Leu Glu Val Ile Glu Asp Cys
Gly Gly Trp Asp Leu Gly Gly Ala Glu Glu Arg Pro Gly Val Ala Ala
Arg Trp Asp Leu Asn Arg Leu Leu Tyr Lys Ala Gln Gly Val Tyr Ser
Ala Ala Ala Leu Phe Ser Leu Thr Val Ser Leu Asp Asp Arg Asn Ser
Ser Arg Tyr Val Ile Arg Ile Asp Gln Asp Gly Leu Thr Leu Pro Glu
Arg Thr Leu Tyr Leu Ala Gln Asp Glu Asp Ser Glu Lys Ile Leu Ala
Ala Tyr Arg Val Phe Met Glu Arg Val Leu Ser Leu Leu Gly Ala Asp

Ala Val Glu Gln Lys Ala Gln Glu Ile Leu Gln Val Glu Gln Gln Leu
Ala Asn Ile Thr Val Ser Glu Tyr Asp Asp Leu Arg Arg Asp Val Ser
Ser Met Tyr Asn Lys Val Thr Leu Gly Gln Leu Gln Lys Ile Thr Pro
His Leu Arg Trp Lys Trp Leu Leu Asp Gln Ile Phe Gln Glu Asp Phe
Ser Glu Glu Glu Glu Val Val Leu Leu Ala Thr Asp Tyr Met Gln Gln
Val Ser Gln Leu Ile Arg Ser Thr Pro His Arg Val Leu His Asn Tyr
Leu Val Trp Arg Val Val Val Val Leu Ser Glu His Leu Ser Pro Pro
Phe Arg Glu Ala Leu His Glu Leu Ala Gln Glu Met Glu Gly Ser Asp
Lys Pro Gln Glu Leu Ala Arg Val Cys Leu Gly Gln Ala Asn Arg His
Phe Gly Met Ala Leu Gly Ala Leu Phe Val His Glu His Phe Ser Ala
Ala Ser Lys Ala Lys Val Gln Gln Leu Val Glu Asp Ile Lys Tyr Ile
Leu Gly Gln Arg Leu Glu Glu Leu Asp Trp Met Asp Ala Glu Thr Arg
Ala Ala Ala Arg Ala Lys Leu Gln Tyr Met Met Val Met Val Gly Tyr
Pro Asp Phe Leu Leu Lys Pro Asp Ala Val Asp Lys Glu Tyr Glu Phe
Glu Val His Glu Lys Thr Tyr Phe Lys Asn Ile Leu Asn Ser Ile Arg
Phe Ser Ile Gln Leu Ser Val Lys Lys Ile Arg Gln Glu Val Asp Lys
Ser Thr Trp Leu Leu Pro Pro Gln Ala Leu Asn Ala Tyr Tyr Leu Pro
Asn Lys Asn Gln Met Val Phe Pro Ala Gly Ile Leu Gln Pro Thr Leu
Tyr Asp Pro Asp Phe Pro Gln Ser Leu Asn Tyr Gly Gly Ile Gly Thr
Ile Ile Gly His Glu Leu Thr His Gly Tyr Asp Asp Trp Gly Gly Gln
Tyr Asp Arg Ser Gly Asn Leu Leu His Trp Trp Thr Glu Ala Ser Tyr
Ser Arg Phe Leu Arg Lys Ala Glu Cys Ile Val Arg Leu Tyr Asp Asn
Phe Thr Val Tyr Asn Gln Arg Val Asn Gly Lys His Thr Leu Gly Glu
Asn Ile Ala Asp Met Gly Gly Leu Lys Leu Ala Tyr His Ala Tyr Gln
Lys Trp Val Arg Glu His Gly Pro Glu His Pro Leu Pro Arg Leu Lys
Tyr Thr His Asp Gln Leu Phe Phe Ile Ala Phe Ala Gln Asn Trp Cys
Ile Lys Arg Arg Ser Gln Ser Ile Tyr Leu Gln Val Leu Thr Asp Lys
His Ala Pro Glu His Tyr Arg Val Leu Gly Ser Val Ser Gln Phe Glu
Glu Phe Gly Arg Ala Phe His Cys Pro Lys Asp Ser Pro Met Asn Pro
Ala His Lys Cys Ser Val Trp (SEQ ID NO:2), which comprises the amino acid sequence of wild type human ECE-3 protein.

As with many enzymes, it is possible to modify many of the amino acids of ECE-3, particularly those which are not found within or in contact with the active site, and still retain substantially the same biological activity as the wild type

protein. Thus this invention includes modified ECE-3 polypeptides which have amino acid deletions, additions, or substitutions but that still retain substantially the same biological activity as ECE-3. It is generally accepted that single amino acid substitutions do not usually alter the biological activity of a protein (see, e.g., *Molecular Biology of the Gene*, Watson *et al.*, 1987, Fourth Ed., The Benjamin/Cummings Publishing Co., Inc., page 226; and Cunningham & Wells, 1989, *Science* 244:1081-1085). Accordingly, the present invention includes polypeptides where one amino acid substitution has been made in SEQ ID NO:2 wherein the polypeptides still retain substantially the same biological activity as ECE-3. The present invention also includes polypeptides where two or more amino acid substitutions have been made in SEQ ID NO:2 wherein the polypeptides still retain substantially the same biological activity as ECE-3. In particular, the present invention includes embodiments where the above-described substitutions are conservative substitutions. In particular, the present invention includes embodiments where the above-described substitutions do not occur within the active site of ECE-3.

One skilled in the art would also recognize that polypeptides that are functional equivalents of ECE-3 and have changes from the ECE-3 amino acid sequence that are small deletions or insertions of amino acids could also be produced by following the same guidelines, (i.e, minimizing the differences in amino acid sequence between ECE-3 and related proteins. Small deletions or insertions are generally in the range of about 1 to 5 amino acids. The effect of such small deletions or insertions on the biological activity of the modified ECE-3 polypeptide can easily be assayed by producing the polypeptide synthetically or by making the required changes in DNA encoding ECE-3 and then expressing the DNA recombinantly and assaying the protein produced by such recombinant expression.

The present invention also includes truncated forms of ECE-3 which contain the region comprising the active site of the enzyme. Such truncated proteins are useful in various assays described herein, for crystallization studies, and for structure-activity-relationship studies.

The present invention also relates to crude or substantially purified subcellular membrane fractions from the recombinant host cells (both prokaryotic and eukaryotic as well as both stably and transiently transformed cells) which contain the nucleic acid molecules of the present invention. These recombinant host cells express ECE-3 or a functional equivalent, which becomes post translationally associated with

an appropriate membrane (such as the cell membrane or the Golgi membrane) in a biologically active fashion. These subcellular membrane fractions will comprise either wild-type or mutant forms of human ECE-3 at levels substantially above endogenous levels and hence will be useful in various assays described throughout this specification. In other words, a specific use for such subcellular membranes involves expression of ECE-3 within the recombinant cell followed by isolation and substantial purification of the membranes away from other cellular components. These substantially purified membranes preparations, which again may be retrieved from a prokaryotic or eukaryotic host cell (including a human recombinant host cell line), will be especially useful in assays to determine the effect of a test substance on ECE-3 catalysis of a big ET precursor to the respective endothelin.

The present invention also relates to isolated nucleic acid molecules which are fusion constructions expressing fusion proteins useful in assays to identify compounds which modulate wild-type vertebrate ECE-3 activity, as well as generating antibodies against human ECE-3. A preferred aspect of this portion of the invention includes, but is not limited to, glutathione S-transferase (GST)-ECE-3 fusion constructs which include, but are not limited to, either the intracellular domain of human ECE-3 as an in-frame fusion at the carboxy terminus of the GST gene or the extracellular and transmembrane ligand binding domain of ECE-3 fused to an GST or immunoglobulin gene by methods known to one of ordinary skill in the art. Recombinant GST-ECE-3 fusion proteins may be expressed in various expression systems, including *Spodoptera frugiperda* (Sf21) insect cells (Invitrogen) using a baculovirus expression vector (pAcG2T, Pharmingen).

Any of a variety of procedures may be used to clone human ECE-3. These methods include, but are not limited to, (1) a RACE PCR cloning technique (Frohman, et al., 1988, *Proc. Natl. Acad. Sci. USA* 85: 8998-9002). 5′ and/or 3′ RACE may be performed to generate a full-length cDNA sequence. This strategy involves using gene-specific oligonucleotide primers for PCR amplification of human ECE-3 cDNA. These gene-specific primers are designed through identification of an expressed sequence tag (EST) nucleotide sequence which has been identified by searching any number of publicly available nucleic acid and protein databases; (2) direct functional expression of the human ECE-3 cDNA following the construction of a human ECE-3-containing cDNA library in an appropriate expression vector system; (3) screening a human ECE-3-containing cDNA library constructed in a bacteriophage

or plasmid shuttle vector with a labeled degenerate oligonucleotide probe designed from the amino acid sequence of the human ECE-3 protein; (4) screening a human ECE-3-containing cDNA library constructed in a bacteriophage or plasmid shuttle vector with a partial cDNA encoding the human ECE-3 protein. This partial cDNA is obtained by the specific PCR amplification of human ECE-3 DNA fragments through the design of degenerate oligonucleotide primers from the amino acid sequence known for other kinases which are related to the human ECE-3 protein; (5) screening a human ECE-3-containing cDNA library constructed in a bacteriophage or plasmid shuttle vector with a partial cDNA or oligonucleotide with homology to a mammalian ECE-3 protein. This strategy may also involve using gene-specific oligonucleotide primers for PCR amplification of human ECE-3 cDNA identified as an EST as described above; or (6) designing 5′ and 3′ gene specific oligonucleotides using SEQ ID NO: 1 as a template so that either the full-length cDNA may be generated by known RACE techniques, or a portion of the coding region may be generated by these same known RACE techniques to generate and isolate a portion of the coding region to use as a probe to screen one of numerous types of cDNA and/or genomic libraries in order to isolate a full-length version of the nucleotide sequence encoding human ECE-3.

It is readily apparent to those skilled in the art that other types of libraries, as well as libraries constructed from other cell types-or species types, may be useful for isolating a human ECE-3-encoding DNA or a human ECE-3 homologue. Other types of libraries include, but are not limited to, cDNA libraries derived from other human cells.

It is readily apparent to those skilled in the art that suitable cDNA libraries may be prepared from cells or cell lines which have ECE-3 activity. The selection of cells or cell lines for use in preparing a cDNA library to isolate a cDNA encoding human ECE-3 may be done by first measuring cell-associated ECE-3 activity using any known assay available for such a purpose.

Preparation of cDNA libraries can be performed by standard techniques well known in the art. Well known cDNA library construction techniques can be found for example, in Sambrook et al., 1989, *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. Complementary DNA libraries may also be obtained from numerous commercial sources, including but not limited to Clontech Laboratories, Inc. and Stratagene.

It is also readily apparent to those skilled in the art that DNA encoding human ECE-3 may also be isolated from a suitable genomic DNA library. Construction of genomic DNA libraries can be performed by standard techniques well known in the art. Well known genomic DNA library construction techniques can be found in Sambrook, et al., *supra*. Genomic clones containing the ECE-3 gene can be obtained from commercially available human PAC or BAC libraries, *e.g.*, from Research Genetics, Huntsville, AL. Alternatively, one may prepare genomic libraries, especially in P1 artificial chromosome vectors, from which genomic clones containing the ECE-3 can be isolated, using probes based upon the ECE-3 nucleotide sequences disclosed herein. Methods of preparing such libraries are known in the art (Ioannou et al.,1994. *Nature Genet.* 6:84-89).

In order to clone the human ECE-3 gene by one of the preferred methods, the amino acid sequence or DNA sequence of human ECE-3 or a homologous protein may be necessary. To accomplish this, the ECE-3 protein or a homologous protein may be purified and partial amino acid sequence determined by automated sequenators. It is not necessary to determine the entire amino acid sequence, but the linear sequence of two regions of 6 to 8 amino acids can be determined for the PCR amplification of a partial human ECE-3 DNA fragment. Once suitable amino acid sequences have been identified, the DNA sequences capable of encoding them are synthesized. Because the genetic code is degenerate, more than one codon may be used to encode a particular amino acid, and therefore, the amino acid sequence can be encoded by any of a set of similar DNA oligonucleotides. Only one member of the set will be identical to the human ECE-3 sequence but others in the set will be capable of hybridizing to human ECE-3 DNA even in the presence of DNA oligonucleotides with mismatches. The mismatched DNA oligonucleotides may still sufficiently hybridize to the human ECE-3 DNA to permit identification and isolation of human ECE-3 encoding DNA. Alternatively, the nucleotide sequence of a region of an expressed sequence may be identified by searching one or more available genomic databases. Gene-specific primers may be used to perform PCR amplification of a cDNA of interest from either a cDNA library or a population of cDNAs. As noted above, the appropriate nucleotide sequence for use in a PCR-based method may be obtained from SEQ ID NO: 1, either for the purpose of isolating overlapping 5′ and 3′ RACE products for generation of a full-length sequence coding for human ECE-3, or to isolate a portion of the nucleotide sequence coding for human

ECE-3 for use as a probe to screen one or more cDNA- or genomic-based libraries to isolate a full-length sequence encoding human ECE-3 or human ECE-3-like proteins.

This invention also includes vectors containing a ECE-3 gene, host cells containing the vectors, and methods of making substantially pure ECE-3 protein comprising the steps of introducing the ECE-3 gene into a host cell, and cultivating the host cell under appropriate conditions such that ECE-3 is produced. The ECE-3 so produced may be harvested from the host cells in conventional ways. Therefore, the present invention also relates to methods of expressing the human ECE-3 protein and biological equivalents disclosed herein, assays employing these gene products, recombinant host cells which comprise DNA constructs which express these proteins, and compounds identified through these assays which act as agonists or antagonists of ECE-3 activity.

The cloned human ECE-3 cDNA obtained through the methods described above may be recombinantly expressed by molecular cloning into an expression vector (such as pcDNA3.neo, pcDNA3.1, pCR2.1, pBlueBacHis2 or pLITMUS28) containing a suitable promoter and other appropriate transcription regulatory elements, and transferred into prokaryotic or eukaryotic host cells to produce recombinant human ECE-3. Techniques for such manipulations can be found described in Sambrook, et al., *supra* , are discussed at length in the Example section and are well known and easily available to the artisan of ordinary skill in the art. Therefore, another aspect of the present invention includes host cells that have been engineered to contain and/or express DNA sequences encoding the ECE-3. An expression vector containing DNA encoding a human ECE-3-like protein may be used for expression of human ECE-3 in a recombinant host cell. Such recombinant host cells can be cultured under suitable conditions to produce ECE-3 or a biologically equivalent form. Recombinant host cells may be prokaryotic or eukaryotic, including but not limited to, bacteria such as *E. coli*, fungal cells such as yeast, mammalian cells including, but not limited to, cell lines of human, bovine, porcine, monkey and rodent origin, and insect cells including but not limited to *Drosophila* and silkworm derived cell lines.

For instance, one insect expression system utilizes *Spodoptera frugiperda* (Sf21) insect cells (Invitrogen) in tandem with a baculovirus expression vector (pAcG2T, Pharmingen). Also, mammalian species which may be suitable and which are commercially available, include but are not limited to, L cells L-M(TK$^-$)

(ATCC CCL 1.3), L cells L-M (ATCC CCL 1.2), Saos-2 (ATCC HTB-85), 293 (ATCC CRL 1573), Raji (ATCC CCL 86), CV-1 (ATCC CCL 70), COS-1 (ATCC CRL 1650), COS-7 (ATCC CRL 1651), CHO-K1 (ATCC CCL 61), 3T3 (ATCC CCL 92), NIH/3T3 (ATCC CRL 1658), HeLa (ATCC CCL 2), C127I (ATCC CRL 1616), BS-C-1 (ATCC CCL 26), MRC-5 (ATCC CCL 171) and CPAE (ATCC CCL 209).

The present invention is also directed to methods for screening for compounds which modulate the expression of DNA or RNA encoding a human ECE-3 protein. Compounds which modulate these activities may be DNA, RNA, peptides, proteins, or non-proteinaceous organic molecules. Compounds may modulate by increasing or attenuating the expression of DNA or RNA encoding human ECE-3, or the function of human ECE-3. Compounds that modulate the expression of DNA or RNA encoding human ECE-3 or the biological function thereof may be detected by a variety of assays. The assay may be a simple "yes/no" assay to determine whether there is a change in expression or function. The assay may be made quantitative by comparing the expression or function of a test sample with the levels of expression or function in a standard sample. Kits containing human ECE-3, antibodies to human ECE-3, or modified human ECE-3 may be prepared by known methods for such uses.

Chinese hamster ovary (CHO) cells are particularly suitable for expression of the ECE-3 protein because these cells express a large number of G-proteins (including ETA and ETB) and contain a minimal amount of ECE activity (Xu et al., 1994, *Cell* 78: 473-485). Thus, it is likely that at least endothelin receptor ETA or ETB will be able to functionally couple the signal generated by interaction with an endothelin peptide generated by the catalytic activity of ECE-3, thus transmitting this signal to downstream effectors, eventually resulting in a measurable change in some assayable component, e.g., cAMP level, expression of a reporter gene, hydrolysis of inositol lipids, or intracellular Ca^{2+} levels. It will also be within the purview of the skilled artisan to generate a CHO cell line which is stably or transiently transfected with a DNA molecule or DNA vector which expresses human ECE-3 and a DNA molecule or DNA vector which expresses an endothelin receptor, including but not limited to a human ETA or human ETB receptor. It will also be within the purview of the skilled artisan to generated a CHO cell line which is stably or transiently transfected with a DNA molecule or DNA vector which expresses human ECE-3, a DNA molecule or DNA vector which expresses an endothelin receptor, including but not limited to a human ETA or human ETB receptor, and a DNA molecule or vector

which expresses either the endothelin prepropolypeptide or the respective big ET polypeptide. While there are numerous variations on this theme which will be evident to the artisan of ordinary skill, including different cell lines, different combinations of DNA molecules or DNA vectors from the endothelin pathway, as well as the stable or transient transfection of the DNA molecule or vector of interest, the construction of the cell lines serves the purpose of conduction cell-based assay to test for modulators of ECE-3 activity or modulators of another step in the endothelin pathway. For example, a cell such as a CHO cell which is transformed with DNA molecules or DNA vectors which express human ECE-3 and a prepropolypeptide, respectively, will be useful in assays to select compounds which modulate ECE-3 activity. These double-transformed or double-transfected cells are allowed to grow for a time sufficient to express human ECE-3 and the endothelin prepropolypeptide wherein the big ET substrate is generated. The transfected cells are harvested and resuspended in a sample assay buffer (plus test compound) and control assay buffer (minus test compound) for an adequate period of time. The conditioned medium from the cells may be directly assayed by known enzyme immunoassay procedures for the presence of mature endothelin peptide. An antagonist of ECE-3 catalyzed conversion of big ET to mature endothelin will result in decreased concentration of processed endothelin versus the control while an agonist of ECE-3 catalyzed conversion to mature endothelin will result in increased concentration of endothelin compared to control levels.

Therefore, the present invention relates to methods of expressing ECE-3 in recombinant systems and of identifying agonists and antagonists of ECE-3. The novel ECE-3 protein of the present invention is suitable for use in an assay procedure for the identification of compounds which modulate the conversion of big ET to ET. Modulating ECE-3 activity, as described herein includes the inhibition or activation of the enzyme and also includes directly or indirectly affecting the normal regulation of the enzymatic activity of ECE-3. Compounds which modulate ECE-3 include agonists, antagonists and compounds which directly or indirectly affect regulation of human ECE-3. When screening compounds in order to identify potential pharmaceuticals that specifically interact with a target protein, it is necessary to ensure that the compounds identified are as specific as possible for the target protein. To do this, it may necessary to screen the compounds against as wide an array as possible of proteins that are similar to the target receptor. Thus, in order to find compounds that

are potential pharmaceuticals that interact with ECE-3, it is necessary not only to ensure that the compounds interact with ECE-3 (the "plus target") and produce the desired pharmacological effect through ECE-3, it is also necessary to determine that the compounds do not interact with proteins B, C, D, etc. (the "minus targets"). In general, as part of a screening program, it is important to have as many minus targets as possible (see Hodgson, 1992, *Bio/Technology* 10:973-980, @ 980). Human ECE-3 proteins and the DNA molecules encoding this protein have the additional utility in that they can be used as "minus targets" in screens designed to identify compounds that specifically interact with other components of the endothelin pathway which effect signal transduction of ET-1, ET-2 and/or ET-3, including but not limited to ECE-1 and ECE-2.

A particular embodiment of the present invention includes a method for determining whether a substance is a potential agonist or antagonist of ECE-3 that comprises:

(a) transfecting cells with an expression vector encoding ECE-3;

(b) allowing the transfected cells to grow for a time sufficient to allow ECE-3 to be expressed;

(c) harvesting the transfected cells and resuspending the cells in the presence of a known, labeled antagonist (including but not limited to phosphoramidon) or agonist of ECE-3 (i.e. ligand for ECE-3) in the presence and in the absence of the substance;

(d) measuring the binding of the labeled ligand for ECE-3 where if the amount of binding of the known ligand is less in the presence of the substance than in the absence of the substance, then the substance is a potential agonist of ECE-3 activity. In a modification of the above-described method, step (c) is modified in that the cells are not harvested and resuspended but rather the radioactively labeled known ligand and the substance are contacted with the cells while the cells are attached to a substratum, e.g., tissue culture plates. The conditions under which step (c) of the method is practiced in this study and other disclosed within this specification are conditions that are typically used in the art for the study of protein-ligand interactions: e.g., physiological pH; salt conditions such as those represented by such commonly used buffers as PBS or in tissue culture media; a temperature of about 4°C to about 55°C. Such ligands need not necessarily be radiolabeled but can also be nonisotopic compounds that can be used to displace bound radiolabeled compounds or

that can be used as activators in functional assays. Compounds identified by the above method are likely to be agonists or antagonists of ECE-3 and may be peptides, proteins, or non-proteinaceous organic molecules.

The present invention also includes a method for determining whether a substance is capable of binding to ECE-3, i.e.,, whether the substance is a potential agonist or an antagonist of ECE-3, where the method comprises:

(a) providing test cells by transfecting cells with an expression vector that directs the expression of ECE-3 in the cells;

(b) exposing the test cells to the substance;

(c) measuring the amount of binding of the substance to ECE-3;

(d) comparing the amount of binding of the substance to ECE-3 in the test cells with the amount of binding of the substance to control cells that have not been transfected with ECE-3;

wherein if the amount of binding of the substance is greater in the test cells as compared to the control cells, the substance is capable of binding to ECE-3. Determining whether the substance is actually an agonist or antagonist can then be accomplished by the use of functional assays such as, e.g., the assay involving the ability of effect conversion of big ET to ET in the presence of ECE-3 and the substance of interest.

The conditions under which step (b) of the method is practiced are conditions that are typically used in the art for the study of protein-ligand interactions: e.g., physiological pH; salt conditions such as those represented by such commonly used buffers as PBS or in tissue culture media; a temperature of about 4°C to about 55°C.

A preferred method for determining whether a substance is capable of binding to endothelin converting enzyme-3 includes the following steps:

(a) providing test cells by transfecting cells with an expression vector that directs the expression of endothelin converting enzyme-3 as set forth as SEQ ID NO:2 in the cells;

(b) exposing the test cells to the substance;

(c) measuring the amount of binding of the substance to the endothelin converting enzyme-3;

(d) comparing the amount of binding of the substance to the endothelin converting enzyme-3 in the test cells with the amount of binding of the

substance to control cells that have not been transfected with the expression vector that directs the expression of endothelin converting enzyme-3 as set forth as SEQ ID NO:2.

A preferred method for determining whether a substance is capable of modulating endothelin converting enzyme-3 activity includes the following steps:

(a) providing test cells by transfecting cells with an expression vector that directs the expression of endothelin converting enzyme-3 as set forth as SEQ ID NO:2;

(b) exposing the test cells to the substance;

(c) measuring the amount of an accumulated intracellular secondary message;

(d) comparing the amount of the secondary message in the test cells in response to the substance with the amount of secondary message in test cells that have not been exposed to the substance.

An especially preferred method measures the amount of accumulated cAMP in a transfected or transformed host cell, such a CHO cell, that endogenously express ETA or ETB and contain minimal endogenous endothelial converting enzyme.

Another preferred screening method of the present invention includes a method for determining whether a substance is capable of modulating endothelin converting enzyme-3 activity which involves some or all of the following steps:

(a) providing test cells by transfecting cells with an expression vector that directs the expression of endothelin converting enzyme-3 as set forth as SEQ ID NO:2;

(b) purifying membrane preparations comprising the endothelin converting enzyme-3;

(c) adding a test substance to the purified membrane preparations of step (b);

(d) incubating the test substance-containing membrane preparation of step (c) with a substrate of endothelin converting enzyme-3;

(e) comparing the product generated from step (d) versus the amount of product generated from a membrane preparation containing the substrate of step (d) without addition of the test substance of step (c).

An especially preferred substrate for use in this enzyme assay is a substrate selected from the group consisting of big ET-1, big ET-2 and big ET-3.

The specificity of binding of compounds showing affinity for ECE-3 is shown by measuring the affinity of the compounds for recombinant cells expressing the protein or for membranes from these cells. Expression of the protein and screening for compounds that bind to ECE-3 or that inhibit the binding of a known, radiolabeled ligand of ECE-3 to these cells, or membranes prepared from these cells, provides an effective method for the rapid selection of compounds with high affinity for ECE-3. Such ligands need not necessarily be radiolabeled but can also be nonisotopic compounds that can be used to displace bound radiolabeled compounds or that can be used as activators in functional assays. Compounds identified by the above method are likely to be agonists or antagonists of ECE-3 and may be peptides, proteins, or non-proteinaceous organic molecules.

An embodiment of the present invention is determining various ligand binding affinities using a labeled ligand in the presence of varying concentration of unlabeled ligands. The activation of the second messenger pathway is determined by measuring the intracellular cAMP elicited by an agonist or antagonist at various concentration. Therefore, it is within the realm of screening for such compounds to use a transactivation assay wherein as host cell is transfected with a reporter construct wherein a reporter gene is fused downstream of a promoter and response element which is positively regulated by a secondary signal such a cAMP. For example, Chen et al. (1995, *Analytical Biochemistry* 226: 349-354) describe a colorimetric assay which utilizes a recombinant cell transfected with an expression vector encoding a G protein coupled receptor with a second expression vector containing a promoter with a cAMP responsive element fused to the LacZ gene. Activity of the overexpressed G-protein coupled receptor is measured as the expression and OD measurement of ß-Gal. Therefore, another aspect of this portion of the invention includes a non-radioactive method for determining whether a substance is a potential agonist or antagonist of ECE-3 that comprises:

(a) stably transfecting or transforming cells with an expression vector encoding ECE-3, wherein these recombinant host cells express biologically active amounts of a G-protein coupled receptor which transmits a biological signal in response to binding with an endothelin;

(b) transiently or stably transfecting the recombinant host cell line of step (a) with an expression vector which comprises a cAMP-inducible promoter fused to a colorimetric gene such a LacZ;

(b) allowing the transfected cells to grow for a time sufficient to allow ECE-3 to be expressed;

(c) harvesting the transfected cells and resuspending the cells in the presence of a known agonist of ECE-3 and/or in both the presence and absence of the test compound;

(d) measuring the binding of the known agonist and test compound to expressed ECE-3 by a colorimetric assay which measures expression off the cAMP-inducible promoter and comparing expression levels in the presence of the known agonist as well as in the presence and absence of the unknown substance so as to determine whether the unknown substance acts as either a potential agonist or antagonist of ECE-3.

It is also possible to transfect or transform the above cells with a construct comprising a third gene wherein this third gene encodes G-protein coupled receptor which interacts with endothelins, such as ETA or ETB. The assays described above can be carried out with cells that have been transiently or stably transfected or stably transformed with expression vectors which encode ECE-3. Transfection is meant to include any method known in the art for introducing ECE-3 into the test cells. For example, transfection includes calcium phosphate or calcium chloride mediated transfection, lipofection, infection with a retroviral construct containing ECE-3, and electroporation. Transformation is meant to encompass a genetic change to the target cell resulting from an incorporation of DNA.

The DNA molecules, RNA molecules, recombinant protein and antibodies of the present invention may be used to screen and measure levels of human ECE-3. The recombinant proteins, DNA molecules, RNA molecules and antibodies lend themselves to the formulation of kits suitable for the detection and typing of human ECE-3. Such a kit would comprise a compartmentalized carrier suitable to hold in close confinement at least one container. The carrier would further comprise reagents such as recombinant ECE-3 or anti-ECE-3 antibodies suitable for detecting human ECE-3. The carrier may also contain a means for detection such as labeled antigen or enzyme substrates or the like.

A variety of mammalian expression vectors may be used to express recombinant human ECE-3 in mammalian cells. Expression vectors are defined herein as DNA sequences that are required for the transcription of cloned DNA and the translation of their mRNAs in an appropriate host. Such vectors can be used to

express eukaryotic DNA in a variety of hosts such as bacteria, blue green algae, plant cells, insect cells and animal cells. Specifically designed vectors allow the shuttling of DNA between hosts such as bacteria-yeast or bacteria-animal cells. An appropriately constructed expression vector should contain: an origin of replication for autonomous replication in host cells, selectable markers, a limited number of useful restriction enzyme sites, a potential for high copy number, and active promoters. A promoter is defined as a DNA sequence that directs RNA polymerase to bind to DNA and initiate RNA synthesis. A strong promoter is one which causes mRNAs to be initiated at high frequency.

Expression vectors may include, but are not limited to, cloning vectors, modified cloning vectors, specifically designed plasmids or viruses. Commercially available mammalian expression vectors which may be suitable for recombinant human ECE-3 expression, include but are not limited to, pcDNA3.neo (Invitrogen), pcDNA3.1 (Invitrogen), pCI-neo (Promega), pLITMUS28, pLITMUS29, pLITMUS38 and pLITMUS39 (New England Bioloabs), pcDNAI, pcDNAIamp (Invitrogen), pcDNA3 (Invitrogen), pMC1neo (Stratagene), pXT1 (Stratagene), pSG5 (Stratagene), EBO-pSV2-neo (ATCC 37593) pBPV-1(8-2) (ATCC 37110), pdBPV-MMTneo(342-12) (ATCC 37224), pRSVgpt (ATCC 37199), pRSVneo (ATCC 37198), pSV2-dhfr (ATCC 37146), pUCTag (ATCC 37460), and lZD35 (ATCC 37565).

Also, a variety of bacterial expression vectors may be used to express recombinant human ECE-3 in bacterial cells. Commercially available bacterial expression vectors which may be suitable for recombinant human ECE-3 expression include, but are not limited to pCR2.1 (Invitrogen), pET11a (Novagen), lambda gt11 (Invitrogen), and pKK223-3 (Pharmacia).

In addition, a variety of fungal cell expression vectors may be used to express recombinant human ECE-3 in fungal cells. Commercially available fungal cell expression vectors which may be suitable for recombinant human ECE-3 expression include but are not limited to pYES2 (Invitrogen) and *Pichia* expression vector (Invitrogen).

Also, a variety of insect cell expression vectors may be used to express recombinant protein in insect cells. Commercially available insect cell expression vectors which may be suitable for recombinant expression of human ECE-3 include

but are not limited to pBlueBacIII and pBlueBacHis2 (Invitrogen). and pAcG2T (Pharmingen).

The assays described above can be carried out with cells that have been transiently or stably transfected with ECE-3. The expression vector may be introduced into host cells via any one of a number of techniques including but not limited to transformation, transfection, protoplast fusion, and electroporation. Transfection is meant to include any method known in the art for introducing ECE-3 into the test cells. For example, transfection includes calcium phosphate or calcium chloride mediated transfection, lipofection, infection with a retroviral construct containing ECE-3, and electroporation. The expression vector-containing cells are individually analyzed to determine whether they produce human ECE-3 protein. Identification of human ECE-3 expressing cells may be done by several means, including but not limited to immunological reactivity with anti-human ECE-3 antibodies, labeled ligand binding, the presence of host cell-associated human ECE-3 activity via the conversion of big ET to ET.

Expression of human ECE-3 DNA may also be performed using *in vitro* produced synthetic mRNA. Synthetic mRNA can be efficiently translated in various cell-free systems, including but not limited to wheat germ extracts and reticulocyte extracts, as well as efficiently translated in cell based systems, including but not limited to microinjection into frog oocytes, with microinjection into frog oocytes being preferred.

To determine the human ECE-3 cDNA sequence(s) that yields optimal levels of human ECE-3, cDNA molecules including but not limited to the following can be constructed: a cDNA fragment containing the full-length open reading frame for human ECE-3 as well as various constructs containing portions of the cDNA encoding only specific domains of the protein or rearranged domains of the protein. All constructs can be designed to contain none, all or portions of the 5′ and/or 3′ untranslated region of a human ECE-3 cDNA. The expression levels and activity of human ECE-3 can be determined following the introduction, both singly and in combination, of these constructs into appropriate host cells. Following determination of the human ECE-3 cDNA cassette yielding optimal expression in transient assays, this ECE-3 cDNA construct is transferred to a variety of expression vectors (including recombinant viruses), including but not limited to those for mammalian cells, plant cells, insect cells, oocytes, bacteria, and yeast cells.

Following expression of human ECE-3 in a host cell, ECE-3 protein may be recovered to provide ECE-3 protein in active form. Several ECE-3 protein purification procedures are available and suitable for use. Recombinant ECE-3 protein may be purified from cell lysates and extracts by various combinations of, or individual application of salt fractionation, ion exchange chromatography, size exclusion chromatography, hydroxylapatite adsorption chromatography and hydrophobic interaction chromatography. In addition, recombinant ECE-3 protein can be separated from other cellular proteins by use of an immunoaffinity column made with monoclonal or polyclonal antibodies specific for full-length ECE-3 protein, or polypeptide fragments of ECE-3 protein.

Polyclonal or monoclonal antibodies may be raised against human ECE-3 or a synthetic peptide (usually from about 9 to about 25 amino acids in length) from a portion of human ECE-3 as disclosed in SEQ ID NO:2. Monospecific antibodies to human ECE-3 are purified from mammalian antisera containing antibodies reactive against human ECE-3 or are prepared as monoclonal antibodies reactive with human ECE-3 using the technique of Kohler and Milstein (1975, *Nature* 256: 495-497). Monospecific antibody as used herein is defined as a single antibody species or multiple antibody species with homogenous binding characteristics for human ECE-3. Homogenous binding as used herein refers to the ability of the antibody species to bind to a specific antigen or epitope, such as those associated with human ECE-3, as described above. Human ECE-3-specific antibodies are raised by immunizing animals such as mice, rats, guinea pigs, rabbits, goats, horses and the like, with an appropriate concentration of human ECE-3 protein or a synthetic peptide generated from a portion of human ECE-3 with or without an immune adjuvant.

Preimmune serum is collected prior to the first immunization. Each animal receives between about 0.1 mg and about 1000 mg of human ECE-3 protein associated with an acceptable immune adjuvant. Such acceptable adjuvants include, but are not limited to, Freund's complete, Freund's incomplete, alum-precipitate, water in oil emulsion containing *Corynebacterium parvum* and tRNA. The initial immunization consists of human ECE-3 protein or peptide fragment thereof in, preferably, Freund's complete adjuvant at multiple sites either subcutaneously (SC), intraperitoneally (IP) or both. Each animal is bled at regular intervals, preferably weekly, to determine antibody titer. The animals may or may not receive booster injections following the initial immunization. Those animals receiving booster

injections are generally given an equal amount of human ECE-3 in Freund's incomplete adjuvant by the same route. Booster injections are given at about three week intervals until maximal titers are obtained. At about 7 days after each booster immunization or about weekly after a single immunization, the animals are bled, the serum collected, and aliquots are stored at about -20°C.

Monoclonal antibodies (mAb) reactive with human ECE-3 are prepared by immunizing inbred mice, preferably Balb/c, with human ECE-3 protein. The mice are immunized by the IP or SC route with about 1 mg to about 100 mg, preferably about 10 mg, of human ECE-3 protein in about 0.5 ml buffer or saline incorporated in an equal volume of an acceptable adjuvant, as discussed above. Freund's complete adjuvant is preferred. The mice receive an initial immunization on day 0 and are rested for about 3 to about 30 weeks. Immunized mice are given one or more booster immunizations of about 1 to about 100 mg of human ECE-3 in a buffer solution such as phosphate buffered saline by the intravenous (IV) route. Lymphocytes, from antibody positive mice, preferably splenic lymphocytes, are obtained by removing spleens from immunized mice by standard procedures known in the art. Hybridoma cells are produced by mixing the splenic lymphocytes with an appropriate fusion partner, preferably myeloma cells, under conditions which will allow the formation of stable hybridomas. Fusion partners may include, but are not limited to: mouse myelomas P3/NS1/Ag 4-1; MPC-11; S-194 and Sp 2/0, with Sp 2/0 being preferred. The antibody producing cells and myeloma cells are fused in polyethylene glycol, about 1000 mol. wt., at concentrations from about 30% to about 50%. Fused hybridoma cells are selected by growth in hypoxanthine, thymidine and aminopterin supplemented Dulbecco's Modified Eagles Medium (DMEM) by procedures known in the art. Supernatant fluids are collected form growth positive wells on about days 14, 18, and 21 and are screened for antibody production by an immunoassay such as solid phase immunoradioassay (SPIRA) using human ECE-3 as the antigen. The culture fluids are also tested in the Ouchterlony precipitation assay to determine the isotype of the mAb. Hybridoma cells from antibody positive wells are cloned by a technique such as the soft agar technique of MacPherson, 1973, Soft Agar Techniques, in *Tissue Culture Methods and Applications*, Kruse and Paterson, Eds., Academic Press.

Monoclonal antibodies are produced *in vivo* by injection of pristine primed Balb/c mice, approximately 0.5 ml per mouse, with about 2 x 10^6 to about 6 x

10^6 hybridoma cells about 4 days after priming. Ascites fluid is collected at approximately 8-12 days after cell transfer and the monoclonal antibodies are purified by techniques known in the art.

In vitro production of anti-human ECE-3 mAb is carried out by growing the hybridoma in DMEM containing about 2% fetal calf serum to obtain sufficient quantities of the specific mAb. The mAb are purified by techniques known in the art.

Antibody titers of ascites or hybridoma culture fluids are determined by various serological or immunological assays which include, but are not limited to, precipitation, passive agglutination, enzyme-linked immunosorbent antibody (ELISA) technique and radioimmunoassay (RIA) techniques. Similar assays are used to detect the presence of human ECE-3 in body fluids or tissue and cell extracts.

It is readily apparent to those skilled in the art that the above described methods for producing monospecific antibodies may be utilized to produce antibodies specific for human ECE-3 peptide fragments, or full-length human ECE-3.

Human ECE-3 antibody affinity columns are made, for example, by adding the antibodies to Affigel-10 (Biorad), a gel support which is pre-activated with N-hydroxysuccinimide esters such that the antibodies form covalent linkages with the agarose gel bead support. The antibodies are then coupled to the gel via amide bonds with the spacer arm. The remaining activated esters are then quenched with 1M ethanolamine HCl (pH 8). The column is washed with water followed by 0.23 M glycine HCl (pH 2.6) to remove any non-conjugated antibody or extraneous protein. The column is then equilibrated in phosphate buffered saline (pH 7.3) and the cell culture supernatants or cell extracts containing full-length human ECE-3 or human ECE-3 protein fragments are slowly passed through the column. The column is then washed with phosphate buffered saline until the optical density (A_{280}) falls to background, then the protein is eluted with 0.23 M glycine-HCl (pH 2.6). The purified human ECE-3 protein is then dialyzed against phosphate buffered saline.

Pharmaceutically useful compositions comprising modulators of human ECE-3 may be formulated according to known methods such as by the admixture of a pharmaceutically acceptable carrier. Examples of such carriers and methods of formulation may be found in Remington's Pharmaceutical Sciences. To form a pharmaceutically acceptable composition suitable for effective administration, such compositions will contain an effective amount of the protein, DNA, RNA,

modified human ECE-3, or either ECE-3 agonists or antagonists including tyrosine kinase activators or inhibitors.

Therapeutic or diagnostic compositions of the invention are administered to an individual in amounts sufficient to treat or diagnose disorders. The effective amount may vary according to a variety of factors such as the individual's condition, weight, sex and age. Other factors include the mode of administration.

The pharmaceutical compositions may be provided to the individual by a variety of routes such as subcutaneous, topical, oral and intramuscular.

The term "chemical derivative" describes a molecule that contains additional chemical moieties which are not normally a part of the base molecule. Such moieties may improve the solubility, half-life, absorption, etc. of the base molecule. Alternatively the moieties may attenuate undesirable side effects of the base molecule or decrease the toxicity of the base molecule. Examples of such moieties are described in a variety of texts, such as Remington's Pharmaceutical Sciences.

Compounds identified according to the methods disclosed herein may be used alone at appropriate dosages. Alternatively, co-administration or sequential administration of other agents may be desirable.

The present invention also has the objective of providing suitable topical, oral, systemic and parenteral pharmaceutical formulations for use in the novel methods of treatment of the present invention. The compositions containing compounds identified according to this invention as the active ingredient can be administered in a wide variety of therapeutic dosage forms in conventional vehicles for administration. For example, the compounds can be administered in such oral dosage forms as tablets, capsules (each including timed release and sustained release formulations), pills, powders, granules, elixirs, tinctures, solutions, suspensions, syrups and emulsions, or by injection. Likewise, they may also be administered in intravenous (both bolus and infusion), intraperitoneal, subcutaneous, topical with or without occlusion, or intramuscular form, all using forms well known to those of ordinary skill in the pharmaceutical arts.

Advantageously, compounds of the present invention may be administered in a single daily dose, or the total daily dosage may be administered in divided doses of two, three or four times daily. Furthermore, compounds for the present invention can be administered in intranasal form via topical use of suitable

intranasal vehicles, or via transdermal routes, using those forms of transdermal skin patches well known to those of ordinary skill in that art. To be administered in the form of a transdermal delivery system, the dosage administration will, of course, be continuous rather than intermittent throughout the dosage regimen.

For combination treatment with more than one active agent, where the active agents are in separate dosage formulations, the active agents can be administered concurrently, or they each can be administered at separately staggered times.

The dosage regimen utilizing the compounds of the present invention is selected in accordance with a variety of factors including type, species, age, weight, sex and medical condition of the patient; the severity of the condition to be treated; the route of administration; the renal, hepatic and cardiovascular function of the patient; and the particular compound thereof employed. A physician or veterinarian of ordinary skill can readily determine and prescribe the effective amount of the drug required to prevent, counter or arrest the progress of the condition. Optimal precision in achieving concentrations of drug within the range that yields efficacy without toxicity requires a regimen based on the kinetics of the drug's availability to target sites. This involves a consideration of the distribution, equilibrium, and elimination of a drug.

The following examples are provided to illustrate the present invention without, however, limiting the same hereto.

EXAMPLE 1

Isolation and Characterization of DNA Molecules Encoding Human ECE-3

There are two previously isolated ECE's, ECE-1 and ECE-2. The EST hsubn_1f12.x00 was found in a GSC database query using a number of full length human zinc metalloprotease protein sequences. A blastn search of the Merck.EST database queried with hsubn_1f12.x00 picked up only one other mouse EST AA387792. A blastn against all of GenBank (Merck.DNA) hit bovine ECE-2 (U27341) with about 90% similarity. A Blastx query against Merck.PROTEIN resulted in finding 85% similarity to human ECE-1 (D43698) unigene class Hs.88611. The full length cDNA to hsubn_1f12.x00 was cloned and sequenced and this 1497 base pair cDNA was reblasted. Blastn against Merck.DNA again hit the bovine ECE-

2 (U27341) complete mRNA sequence with 90% identity. The second highest hit was to an EST entered 10 April 1998 to a human mRNA from KIAA0604 (ab011176) at 89% similarity. Blastx also hit the KIAA0604 sequence in the protein form. This human KIAA0604 protein sequence is 89% similar to the bovine ECE-1 sequence and 74% similar to human ECE-1. This indicates that KIAA0604 is most likely an ortholog of bovine ECE-2 and a paralog of human ECE-1. Human ECE-1 protein is also 89% similar to bovine ECE-2. Since the mouse hsubn_1f12.x00 sequence is 89% similar to both bovine ECE-2 and to human KIAA0604, it probably represent the mouse ortholog of the ECE-2 gene.

In a further search of the public database, an EST corresponding to a third member of this family was found, and a full-length sequence was isolated by EST sequencing and RACE (Rapid Amplification of cDNA Ends). A unigene search looking for endothelin converting enzyme hit index class Hs.26880 which consists of 3 ESTs and is annotated as being 44% similar to the rat ECE-1. The three ESTs represent 2 Image clones, which have the following designation and nucleotide sequence:

1. AA523527:
EST Id: 1168146; EST name: ni45g12.s1;
GenBank Acc.#: AA523527; GenBank gi: 2264239;
Clone Id: IMAGE:979846 (3') from NCI;

CTTTGCAGCA AGAACACAGC GAAGGTGGGG CCCGTACAAT CCAGCCTGGC AGAGGGTCTG
GCCCCCTTAG AGCAGAATCT GGGGACCCCA GTATATTTCC CTCACAGCCC CCCAAAGTCC
AGCCTCACCC TGCTCCAGGC CCCTCCTGAA GTGAGGGGCA GCAGGGGGAC CGGGTCCTGG
AGGGGCTGGA AGGCAGGTGG TGCCCAGAGC GGGGCTGGCA CCGGGTGCAT GCCTGCCCCG
GTAGCCAGCA GGAGGTGATT CGTGCGGGGG CAGTGGGGGC GTGCAGGCGG GCAGCAAGGC
TCACCACACG GAACACTTGT GGGCAGGGTT CATGGGTGAG TCCTTGGGAC AGTGGAAAGC
CCGGCCAAAC TCCTCAAACT GNGACACACT GCCCAGCACC CTGTAGTGCT CAGGGGCATG
CTTGTCAGTC AGCAACTGCA GGTAGATGGA CTGCGACCGC GCTTGATGCA CAGTTTTGGG
AAAAGGCATT AAAAG (SEQ ID NO:3);

2. R61440:
EST Id: 238195; EST name: yh15h02.r1;
GenBank Acc.#: R61440; GenBank gi: 832135;
GDB Id: 410527;

Clone Id: 37986 (5');

GGTCGGCTAC CCGGACTTCC TGCTGAAACC CGATGCTGTG GACAAGGAGT ATGAGTTTGA
GGTCCATGAG AAGACCTACT TCAAGAACAT CTTGAACAGC ATCCGCTTCA GCATCCAGCT
CTCAGTTAAG AAGATTCGGC AGGAGGTGGA CAAGTCCACG TGGCTGCTCC CCCCACAGGC
GCTCAATGCC TACTATCTAC CCAACAAGAA CCAGATGGTG TTCCCCGCGG GCATCCTGCA
GCCCACCCTG TACGACCCTG ACTTCCCACA GTCTCTCAAC TACGGGGGCA TCGGCACCAT
CATTGGACAT GAGCTGACCC ACGGCTACGA CGGACTGGGG GGGCCAGTAT GACCGCTCAG
GGAACCTGCT TGCACTGGTG GGACGGAGGC TTCCTTACAG CCGNTTTCCT GCGAAAGGCT
GAGTGCATCG TTCCCTNTTT TATGGACAAC TTTCAATGTN TTACAACCAG GCGGTGAACG
GGAAACACAN GTTTGGGAGA ACATCGCAGT ATGGGGCGGN CTTAAGTTGG CTTACCACGC
TATTAGAGTT GGTTNCGGGA NGGCCCCAGG AGCACCATTT CCCGGTTAAA TACANACTGA
ACCAGT (SEQ ID NO:4); and,

3. R61395:
EST Id: 238150; EST name: yh15h02.s1;
GenBank Acc. #: R61395; GenBank gi: 832090;
GDB Id: 410527
Clone Id: 37986 (3')

TTTTTTTTTT TAACAGTGCA GTGATTTATT GACCAGACTT TGCAGCAAGA ACACAGCGAA
GGTGGGGCCC GTACAATCCA GCCTGGCAGA GGGTCTGGCC CCCTTAGAGC AGAATCTGGG
GACCCCAGTA TATTTCCCTC ACAGCCCCCC AAAGTCCAGC CTCACCCTGC TCCAGGCCCC
TCCTGAAGTG AGGGGCAGCA GGGGGACCGG AGTCCTGGAG GGGCTGGAAG GCAGGTGGTG
CCCAGAGCGG GGCTGGCACC GGGTGCATGC CTGCCCCGGT AGCCAGCAGG AGGTNATTCG
TNCGGGGGCA GTNGGGGCNT GCAGGCGGGC ANCAGGNTTC ACCACACGGA AC
(SEQ ID NO:5).

R61440 is a 5' EST 606 bp long which hits KIAA0604 with 60% similarity over only a portion of the full length EST (167/276 bp). No similar ESTs were found during this search of the public database. The two 3' sequences are identical to each other and to no other EST. The above ESTs may also be viewed at the National Center for Biotechnology Information (NCBI) homepage.

The full-length ECE-3 cDNA sequence was obtained by 5' RACE-PCR. First, the IMAGE clone 37986, from which the ESTs yh15h02.r1 (GenBank Acc.#: R61440) and yh15h02.s1 (GenBank Acc. #: R61395) were derived, and the IMAGE clone 979846 (GenBank Acc. #: AA523527) were obtained from Research Genetics and resequenced to correct errors in the EST sequences. This sequence was compared with those of the other ECE family members (ECE-1 and ECE-2) in the public databases, and it was determined that the 3' terminus of the gene was present in the IMAGE clone 37986 and IMAGE clone 979846 sequence. The 5' sequence was incomplete, however, and was obtained by 5' RACE. Two rounds of 5' RACE PCR were required to obtain the full length ECE-3 sequence; in both cases nested PCR reactions were carried out. The first round of PCR utilized gene specific primers 1-4 below (SEQ ID NO:6-9), based on the sequence of IMAGE clones 979846 and 37986 (979846 provided 244bp of 5' sequence that was not present in the public databases), using the Advantage cDNA PCR Kit (Clontech#K1905-1), and human fetal brain Marathon-Ready cDNA (Clontech #7402-1). Adaptor primers AP-1 and AP-2 were supplied by the manufacturer (Clontech). The first round RACE-PCR products were sequenced and gave 930 bp of additional ECE-3 sequence. Based on this sequence, further gene specific RACE-PCR primers (5-7 below, SEQ ID NO:10-12) were synthesized and used to amplify further 5' sequence, using the human ovary Marathon-Ready cDNA (Clontech #7417-1), and the Advantage-GC cDNA PCR Kit (Clontech #K1907-1). This produced a cDNA of 2894 nucleotides (see SEQ ID NO:1), which corresponded to the size of the band on the northern blot (see Example 2), and had a clear Kozak consensus start site (GGCGCCATGG, nucleotides 206-215 of SEQ ID NO:1), indicating that this sequence does indeed represent the full length ECE-3 gene.

1. 5'-GAAGTCAGGGTCGTACAGGGTG-3' (SEQ ID NO:6)
2. 5'-CTTGTTGGGTAGATAGTAGGC-3' (SEQ ID NO:7)
3. 5'-CTGGATGCTGAAGCGGATGCTGTTC-3' (SEQ ID NO:8)
4. 5'-CCTCAAACTCATACTCCTTGTCCAC-3' (SEQ ID NO:9)
5. 5'-CCTTGTTGTACATGGAGCTGACATC-3' (SEQ ID NO:10)
6. 5'-CCTTCTGTTCCACAGCGTCTGCAC-3' (SEQ ID NO:11)
7. 5'-GGTGAGCCCATCCTGGTCAATGCG-3' (SEQ ID NO:12).

PCR conditions were as follows: 94^{0}C for 1min, 5 cycles of 94^{0}C 10 sec, 72^{0}C 4 min; 5 cycles of 94^{0}C 10 sec, 70^{0}C 4 min; 20 cycles for 94^{0}C 10 sec, 68^{0}C 4 min and 68^{0}C for 10 min. PCR products were then cloned using the Zero Blunt PCR Cloning Kit (Invitrogen #K2700-20) and sequenced.

Once the full-length human ECE-3 cDNA sequence was obtained, a full-length human ECE-3 cDNA clone was then produced by PCR from both human fetal brain Marathon-Ready cDNA (Clontech #7402-1) and human ovary Marathon-Ready cDNA (Clontech #7417-1). Sequencing of each product gave the same sequence, showing that the ECE-3 mRNA is the same in both tissues. Two pairs of primers were used:

1. forward: 5'-GCTCGGCTGCGCTGCGGCTCAG-3' (SEQ ID NO:13)
   reverse: 5'-GGTCCTGGAGGGGCTGGAAGG-3' (SEQ ID NO:14)
2. forward: 5'-CATCCCGTAGCCCAGGTGGC-3' (SEQ ID NO:15)
   reverse: 5'-GCCAGCAGGAGGTGATTCGTGCG-3' (SEQ ID NO:16)

PCR conditions were as follows: 94^{0}C for 1 min, 32 cycles for 94^{0}C 12 sec and 68^{0}C 4 min, then 68^{0}C for 10 min. This primary PCR amplification was subjected to nested PCR with primer pair 2 (SEQ ID NOs: 15 and 16). PCR products were subcloned and confirmed by DNA sequencing.

Radiation hybrid mapping was done with the GeneBridge4 panel (Gyapay et al., 1996) consisting of 93 radiation hybrid clones (Research Genetics). Primers had sequences:
forward: 5'-GTCCCCCTGCTGCCCCTCACTTCAG-3' (SEQ ID NO:17); and
reverse: 5'-CCAGACTTTGCAGCAAGAACACAGC-3' (SEQ ID NO:18), and produced the expected band of 175 bp. AmpliTaq Gold (Perkin Elmer) was performed with cycling parameters: 94°C for 1 min; 94°C for 20 sec, 70°C for 2 min, 32 cycles; 70°C for 5 min. Results were submitted to the Whitehead Institute Genome Center server, http://www-genome.wi.mit.edu/cgi-bin/contig/rhmapper.pl, and gave similar results. ECE-3 mapped to chromosome 2q, between markers CHLC.GATA12H10.14 and D2S331, 10.65 cR from CHLC.GATA12H10.14 (lod>15).

Therefore, the exemplified cDNA molecule is 2894 base pairs, with an open reading frame of 2325 base pairs, corresponding to a protein of 775 amino acids. Radiation hybrid mapping assigns this gene to chromosome 2q37, in a region that has not been linked to any human disease that might logically be due to ECE mutation.

EXAMPLE 2

Northern Analysis of Human ECE-3 Expression

Four blots were used for Northern analysis: Human Brain Multiple Tissue Northern (MTN) Blot II (#7755-1); Human Brain Multiple Tissue Northern (MTN) Blot III (#7750-1); Human Multiple Tissue Northern (MTN) Blot (#7760-1); Human Multiple Tissue Northern (MTN) Blot II (#7759-1). A DNA hybridization probe was generated from cDNA sources as disclosed in Example 1 using the following primers:

Forward: 5' CTTCCTGCTGAAACCCGATGC 3' (SEQ ID NO:19);

Reverse: 5' CCAGACTTTGCAGCAAGAACACAGC 3' (SEQ ID NO:20)

The DNA sequence of the probe is as follows:

CTTCCTGCTGAAACCCGATGCTGTGGACAAGGAGTATGAGTTTGAGGTCCATGAGAAGACCTACTTCAA
GAACATCTTGAACAGCATCCGCTTCAGCATCCAGCTCTCAGTTAAGAAGATTCGGCAGGAGGTGGACAA
GTCCACGTGGCTGCTCCCCCCACAGGCGCTCAATGCCTACTATCTACCCAACAAGAACCAGATGGTGTT
CCCCGCGGGCATCCTGCAGCCCACCCTGTACGACCCTGACTTCCCACAGTCTCTCAACTACGGGGGCAT
CGGCACCATCATTGGACATGAGCTGACCCACGGCTACGACGACTGGGGGGGCCAGTATGACCGCTCAGG
GAACCTGCTGCACTGGTGGACGGAGGCCTCCTACAGCCGCTTCCTGCGAAAGGCTGAGTGCATCGTCCG
TCTCTATGACAACTTCACTGTCTACAACCAGCGGGTGAACGGGAAACACACGCTTGGGGAGAACATCGC
AGATATGGGCGGCCTCAAGCTGGCCTACCACGCCTATCAGAAGTGGGTGCGGGAGCACGGCCCAGAGCA
CCCACTTCCCCGGCTCAAGTACACACATGACCAGCTCTTCTTCATTGCCTTTGCCCAGAACTGGTGCAT
CAAGCGGCGGTCGCAGTCCATCTACCTGCAGGTGCTGACTGACAAGCATGCCCCTGAGCACTACAGGGT
GCTGGGCAGTGTGTCCCAGTTTGAGGAGTTTGGCCGGGCTTTCCACTGTCCCAAGGACTCACCCATGAA
CCCTGCCCACAAGTGTTCCGTGTGGTGAGCCTGGCTGCCCGCCTGCACGCCCCCACTGCCCCCGCACGA
ATCACCTCCTGCTGGCTACCGGGGCAGGCATGCACCCGGTGCCAGCCCCGCTCTGGGCACCACCTGCCT
TCCAGCCCCTCCAGGACCCGGTCCCCCTGCTGCCCCTCACTTCAGGAGGGGCCTGGAGCAGGGTGAGGC
TGGACTTTGGGGGGCTGTGAGGGAAATATACTGGGGTCCCCAGATTCTGCTCTAAGGGGGCCAGACCCT
CTGCCAGGCTGGATTGTACGGGCCCCACCTTCGCTGTGTTCTTGCTGCAAAGTCTGG

(SEQ ID NO:21).

Twenty five to fifty nanograms of probe was labeled with 5 ul of [α-^{32}P] dCTP (3000Ci/mmol, 10mCi/ml) with a DECAprime II DNA Labeling Kit (Ambion#1455). CENTRI-SEP Columns from Princeton Separations (#CS-900) was used for probe purification. The blots were prehybridized in ExpressHyb solution (Clontech #8015-

1) at 68°C for 1 hour, and hybridized at 68°C for 2 hours. Wash conditions were 2xSSC, 0.05%SDS at room temperature for 30 minutes and twice in 0.1xSSC, 0.1%SDS at 50°C for 20 minutes per wash. The blot was exposed overnight on a PhosphorImager (Storm, Molecular Dynamics) at room temperature and analyzed using the Molecular Dynamics ImageQuant software. Northern analysis showed a single band of approximately 3.0 kb, expressed at a high level in CNS (medulla oblongata) and non-CNS (ovary); medium level in CNS (caudate nucleus, putamen, thalamus, substantia nigra, spinal cord) and non-CNS (testis) and low level in CNS (amygalala, hippocampus, subthalamic nucleus, cerebral cortex, occipital pole, frontal lobe, temporal lobe, cerebellum, corpus callosum) and non-CNS (thymus, prostate, skeletal muscle, kidney, pancreas, heart), as shown in Figure 3A-D.

WHAT IS CLAIMED IS:

1. A purified nucleic acid molecule encoding a human endothelin converting enzyme-3 protein which comprises the nucleotide sequence:

GGCGGCGGGC GCTGGGAGAC ACCGGACGCC CGCTCGGCTG CGCTGCGGCT CAGGCCCCCG
CYCGGGCCCG ACCCGCTCGG TCACCGCCGG CTCGGGCGCG CACCTGCCGG CTGCGGCCCC
AGGGCCATGC GGAGGCCCAC GAGGAGGCCG GCGGCCACGC GCATCCCGTA GCCCAGGTGG
CCCAGGTCTG CACCGCGGCG GCCTCGGCGC CATGGAGCCC CCGTATTCGC TGACGGCGCA
CTACGATGAG TTCCAAGAGG TCAAGTACGT GAGCCGCTGC GGCGCGGGGG GCGCGCGCGG
GGCCTCCCTG CCCCCGGGCT TCCCGTTGGG CGCTGCCCGC AGCGCCACCG GGGCCCGGTC
CGGGCTGCCG CGCTGGAACC GGCGCGAGGT GTGCCTGCTG TCGGGGCTGG TGTTCGCCGC
CGGCCTCTGC GCCATTCTGG CGGCTATGCT GGCCCTCAAG TACCTGGGCC CGGTCGCGGC
CGGCGGCGGC GCCTGTCCCG AGGGCTGCCC TGAGCGCAAG GCCTTCGCGC GCGCCGCTCG
CTTCCTGGCC GCCAACCTGG ACGCCAGCAT CGACCCATGC CAGGACTTCT ACTCGTTCGC
CTGCGGCGGT TGGCTGCGGC GCCACGCCAT CCCCGACGAC AAGCTCACCT ATGGCACCAT
CGCGGCCATC GGCGAGCAAA ACGAGGAGCG CCTACGGCGC CTGCTGGCGC GGCCCGGGGG
TGGGCCTGGC GGCGCGGCCC AGCGCAAGGT GCGCGCCTTC TTCCGCTCGT GCCTCGACAT
GCGCGAGATC GAGCGACTGG GCCCGCGACC CATGCTAGAG GTCATCGAGG ACTGCGGGGG
CTGGGACCTG GGCGGCGCGG AGGAGCGTCC GGGGGTCGCG GCGCGATGGG ACCTCAACCG
GCTGCTGTAC AAGGCGCAGG GCGTGTACAG CGCCGCCGCG CTCTTCTCGC TCACGGTCAG
CCTGGACGAC AGGAACTCCT CGCGCTACGT CATCCGCATT GACCAGGATG GGCTCACCCT
GCCAGAGAGG ACCCTGTACC TCGCTCAGGA TGAGGACAGT GAGAAGATCC TGGCAGCATA
CAGGGTGTTC ATGGAGCGAG TGCTCAGCCT CCTGGGTGCA GACGCTGTGG AACAGAAGGC
CCAAGAGATC CTGCAAGTGG AGCAGCAGCT GGCCAACATC ACTGTGTCAG AGTATGACGA
CCTACGGCGA GATGTCAGCT CCATGTACAA CAAGGTGACG CTGGGGCAGC TGCAGAAGAT
CACCCCCCAC TTGCGGTGGA AGTGGCTGCT AGACCAGATC TTCCAGGAGG ACTTCTCAGA
GGAAGAGGAG GTGGTGCTGC TGGCGACAGA CTACATGCAG CAGGTGTCGC AGCTCATCCG
CTCCACACCC CACCGGGTCC TGCACAACTA CCTGGTGTGG CGCGTGGTGG TGGTCCTGAG
TGAACACCTG TCCCCGCCAT TCCGTGAGGC ACTGCACGAG CTGGCACAGG AGATGGAGGG
CAGCGACAAG CCACAGGAGC TGGCCCGGGT CTGCTTGGGC CAGGCCAATC GCCACTTTGG
CATGGCGCTT GGCGCCCTCT TTGTACATGA GCACTTCTCA GCTGCCAGCA AAGCCAAGGT
GCAGCAGCTA GTGGAAGACA TCAAGTACAT CCTGGGCCAG CGCCTGGAGG AGCTGGACTG
GATGGACGCC GAGACCAGGG CTGCTGCTCG GGCCAAGCTC CAGTACATGA TGGTGATGGT
CGGCTACCCG GACTTCCTGC TGAAACCCGA TGCTGTGGAC AAGGAGTATG AGTTTGAGGT

CCATGAGAAG ACCTACTTCA AGAACATCTT GAACAGCATC CGCTTCAGCA TCCAGCTCTC
AGTTAAGAAG ATTCGGCAGG AGGTGGACAA GTCCACGTGG CTGCTCCCCC CACAGGCGCT
CAATGCCTAC TATCTACCCA ACAAGAACCA GATGGTGTTC CCCGCGGGCA TCCTGCAGCC
CACCCTGTAC GACCCTGACT TCCCACAGTC TCTCAACTAC GGGGGCATCG GCACCATCAT
TGGACATGAG CTGACCCACG GCTACGACGA CTGGGGGGGC CAGTATGACC GCTCAGGGAA
CCTGCTGCAC TGGTGGACGG AGGCCTCCTA CAGCCGCTTC CTGCGAAAGG CTGAGTGCAT
CGTCCGTCTC TATGACAACT TCACTGTCTA CAACCAGCGG GTGAACGGGA AACACACGCT
TGGGGAGAAC ATCGCAGATA TGGGCGGCCT CAAGCTGGCC TACCACGCCT ATCAGAAGTG
GGTGCGGGAG CACGGCCCAG AGCACCCACT TCCCCGGCTC AAGTACACAC ATGACCAGCT
CTTCTTCATT GCCTTTGCCC AGAACTGGTG CATCAAGCGG CGGTCGCAGT CCATCTACCT
GCAGGTGCTG ACTGACAAGC ATGCCCCTGA GCACTACAGG GTGCTGGGCA GTGTGTCCCA
GTTTGAGGAG TTTGGCCGGG CTTTCCACTG TCCCAAGGAC TCACCCATGA ACCCTGCCCA
CAAGTGTTCC GTGTGGTGAG CCTGGCTGCC CGCCTGCACG CCCCCACTGC CCCCGCACGA
ATCACCTCCT GCTGGCTACC GGGGCAGGCA TGCACCCGGT GCCAGCCCCG CTCTGGGCAC
CACCTGCCTT CCAGCCCCTC CAGGACCCGG TCCCCCTGCT GCCCCTCACT TCAGGAGGGG
CCTGGAGCAG GGTGAGGCTG GACTTTGGGG GGCTGTGAGG GAAATATACT GGGGTCCCCA
GATTCTGCTC TAAGGGGGCC AGACCCTCTG CCAGGCTGGA TTGTACGGGC CCCACCTTCG
CTGTGTTCTT GCTGCAAAGT CTGGTCAATA AATCACTGCA CTGTTAAAAA AAAAAAAAAA
AAAAAATTCC TGCG (SEQ ID NO:1).

2. A purified DNA molecule encoding human endothelin converting enzyme-3 protein wherein said DNA molecule encodes a protein comprising the amino acid sequence:

Met Glu Pro Pro Tyr Ser Leu Thr Ala His Tyr Asp Glu Phe Gln Glu
Val Lys Tyr Val Ser Arg Cys Gly Ala Gly Gly Ala Arg Gly Ala Ser
Leu Pro Pro Gly Phe Pro Leu Gly Ala Ala Arg Ser Ala Thr Gly Ala
Arg Ser Gly Leu Pro Arg Trp Asn Arg Arg Glu Val Cys Leu Leu Ser
Gly Leu Val Phe Ala Ala Gly Leu Cys Ala Ile Leu Ala Ala Met Leu
Ala Leu Lys Tyr Leu Gly Pro Val Ala Ala Gly Gly Gly Ala Cys Pro
Glu Gly Cys Pro Glu Arg Lys Ala Phe Ala Arg Ala Ala Arg Phe Leu
Ala Ala Asn Leu Asp Ala Ser Ile Asp Pro Cys Gln Asp Phe Tyr Ser
Phe Ala Cys Gly Gly Trp Leu Arg Arg His Ala Ile Pro Asp Asp Lys
Leu Thr Tyr Gly Thr Ile Ala Ala Ile Gly Glu Gln Asn Glu Glu Arg
Leu Arg Arg Leu Leu Ala Arg Pro Gly Gly Gly Pro Gly Gly Ala Ala

Gln Arg Lys Val Arg Ala Phe Phe Arg Ser Cys Leu Asp Met Arg Glu
Ile Glu Arg Leu Gly Pro Arg Pro Met Leu Glu Val Ile Glu Asp Cys
Gly Gly Trp Asp Leu Gly Gly Ala Glu Glu Arg Pro Gly Val Ala Ala
Arg Trp Asp Leu Asn Arg Leu Leu Tyr Lys Ala Gln Gly Val Tyr Ser
Ala Ala Ala Leu Phe Ser Leu Thr Val Ser Leu Asp Asp Arg Asn Ser
Ser Arg Tyr Val Ile Arg Ile Asp Gln Asp Gly Leu Thr Leu Pro Glu
Arg Thr Leu Tyr Leu Ala Gln Asp Glu Asp Ser Glu Lys Ile Leu Ala
Ala Tyr Arg Val Phe Met Glu Arg Val Leu Ser Leu Leu Gly Ala Asp
Ala Val Glu Gln Lys Ala Gln Glu Ile Leu Gln Val Glu Gln Gln Leu
Ala Asn Ile Thr Val Ser Glu Tyr Asp Asp Leu Arg Arg Asp Val Ser
Ser Met Tyr Asn Lys Val Thr Leu Gly Gln Leu Gln Lys Ile Thr Pro
His Leu Arg Trp Lys Trp Leu Leu Asp Gln Ile Phe Gln Glu Asp Phe
Ser Glu Glu Glu Glu Val Val Leu Leu Ala Thr Asp Tyr Met Gln Gln
Val Ser Gln Leu Ile Arg Ser Thr Pro His Arg Val Leu His Asn Tyr
Leu Val Trp Arg Val Val Val Val Leu Ser Glu His Leu Ser Pro Pro
Phe Arg Glu Ala Leu His Glu Leu Ala Gln Glu Met Glu Gly Ser Asp
Lys Pro Gln Glu Leu Ala Arg Val Cys Leu Gly Gln Ala Asn Arg His
Phe Gly Met Ala Leu Gly Ala Leu Phe Val His Glu His Phe Ser Ala
Ala Ser Lys Ala Lys Val Gln Gln Leu Val Glu Asp Ile Lys Tyr Ile
Leu Gly Gln Arg Leu Glu Glu Leu Asp Trp Met Asp Ala Glu Thr Arg
Ala Ala Ala Arg Ala Lys Leu Gln Tyr Met Met Val Met Val Gly Tyr
Pro Asp Phe Leu Leu Lys Pro Asp Ala Val Asp Lys Glu Tyr Glu Phe
Glu Val His Glu Lys Thr Tyr Phe Lys Asn Ile Leu Asn Ser Ile Arg
Phe Ser Ile Gln Leu Ser Val Lys Lys Ile Arg Gln Glu Val Asp Lys
Ser Thr Trp Leu Leu Pro Pro Gln Ala Leu Asn Ala Tyr Tyr Leu Pro
Asn Lys Asn Gln Met Val Phe Pro Ala Gly Ile Leu Gln Pro Thr Leu
Tyr Asp Pro Asp Phe Pro Gln Ser Leu Asn Tyr Gly Gly Ile Gly Thr
Ile Ile Gly His Glu Leu Thr His Gly Tyr Asp Asp Trp Gly Gly Gln
Tyr Asp Arg Ser Gly Asn Leu Leu His Trp Trp Thr Glu Ala Ser Tyr
Ser Arg Phe Leu Arg Lys Ala Glu Cys Ile Val Arg Leu Tyr Asp Asn
Phe Thr Val Tyr Asn Gln Arg Val Asn Gly Lys His Thr Leu Gly Glu
Asn Ile Ala Asp Met Gly Gly Leu Lys Leu Ala Tyr His Ala Tyr Gln
Lys Trp Val Arg Glu His Gly Pro Glu His Pro Leu Pro Arg Leu Lys
Tyr Thr His Asp Gln Leu Phe Phe Ile Ala Phe Ala Gln Asn Trp Cys

Ile Lys Arg Arg Ser Gln Ser Ile Tyr Leu Gln Val Leu Thr Asp Lys
His Ala Pro Glu His Tyr Arg Val Leu Gly Ser Val Ser Gln Phe Glu
Glu Phe Gly Arg Ala Phe His Cys Pro Lys Asp Ser Pro Met Asn Pro
Ala His Lys Cys Ser Val Trp (SEQ ID NO:2).

3. An expression vector for the expression of an endothelin converting enzyme-3 in a recombinant host cell wherein said expression vector comprises a DNA molecule which encodes the amino acid sequence of claim 2.

4. An expression vector of claim 3 which is a eukaryotic expression vector.

5. An expression vector of claim 3 which is a prokaryotic expression vector.

6. A host cell which expresses a recombinant endothelin converting enzyme-3 protein wherein said host cell contains the expression vector of claim 3.

7. A host cell which expresses a recombinant endothelin converting enzyme-3 wherein said host cell contains the expression vector of claim 4.

8. A host cell which expresses a recombinant endothelin converting enzyme-3 protein wherein said host cell contains the expression vector of claim 5.

9. A subcellular membrane fraction obtained from the host cell of claim 6 which contains recombinant endothelin converting enzyme-3 .

10. A subcellular membrane fraction obtained from the host cell of claim 7 which contains recombinant endothelin converting enzyme-3.

11. A subcellular membrane fraction obtained from the host cell of claim 8 which contains recombinant endothelin converting enzyme-3.

12. A purified DNA molecule encoding endothelin converting enzyme-3 which consists of the nucleotide sequence:

GGCGGCGGGC GCTGGGAGAC ACCGGACGCC CGCTCGGCTG CGCTGCGGCT CAGGCCCCCG
CYCGGGCCCG ACCCGCTCGG TCACCGCCGG CTCGGGCGCG CACCTGCCGG CTGCGGCCCC
AGGGCCATGC GGAGGCCCAC GAGGAGGCCG GCGGCCACGC GCATCCCGTA GCCCAGGTGG
CCCAGGTCTG CACCGCGGCG GCCTCGGCGC CATGGAGCCC CCGTATTCGC TGACGGCGCA
CTACGATGAG TTCCAAGAGG TCAAGTACGT GAGCCGCTGC GGCGCGGGGG GCGCGCGCGG
GGCCTCCCTG CCCCCGGGCT TCCCGTTGGG CGCTGCCCGC AGCGCCACCG GGGCCCGGTC
CGGGCTGCCG CGCTGGAACC GGCGCGAGGT GTGCCTGCTG TCGGGGCTGG TGTTCGCCGC
CGGCCTCTGC GCCATTCTGG CGGCTATGCT GGCCCTCAAG TACCTGGGCC CGGTCGCGGC
CGGCGGCGGC GCCTGTCCCG AGGGCTGCCC TGAGCGCAAG GCCTTCGCGC GCGCCGCTCG
CTTCCTGGCC GCCAACCTGG ACGCCAGCAT CGACCCATGC CAGGACTTCT ACTCGTTCGC
CTGCGGCGGT TGGCTGCGGC GCCACGCCAT CCCCGACGAC AAGCTCACCT ATGGCACCAT
CGCGGCCATC GGCGAGCAAA ACGAGGAGCG CCTACGGCGC CTGCTGGCGC GGCCCGGGGG
TGGGCCTGGC GGCGCGGCCC AGCGCAAGGT GCGCGCCTTC TTCCGCTCGT GCCTCGACAT
GCGCGAGATC GAGCGACTGG GCCCGCGACC CATGCTAGAG GTCATCGAGG ACTGCGGGGG
CTGGGACCTG GGCGGCGCGG AGGAGCGTCC GGGGGTCGCG GCGCGATGGG ACCTCAACCG
GCTGCTGTAC AAGGCGCAGG GCGTGTACAG CGCCGCCGCG CTCTTCTCGC TCACGGTCAG
CCTGGACGAC AGGAACTCCT CGCGCTACGT CATCCGCATT GACCAGGATG GGCTCACCCT
GCCAGAGAGG ACCCTGTACC TCGCTCAGGA TGAGGACAGT GAGAAGATCC TGGCAGCATA
CAGGGTGTTC ATGGAGCGAG TGCTCAGCCT CCTGGGTGCA GACGCTGTGG AACAGAAGGC
CCAAGAGATC CTGCAAGTGG AGCAGCAGCT GGCCAACATC ACTGTGTCAG AGTATGACGA
CCTACGGCGA GATGTCAGCT CCATGTACAA CAAGGTGACG CTGGGGCAGC TGCAGAAGAT
CACCCCCCAC TTGCGGTGGA AGTGGCTGCT AGACCAGATC TTCCAGGAGG ACTTCTCAGA
GGAAGAGGAG GTGGTGCTGC TGGCGACAGA CTACATGCAG CAGGTGTCGC AGCTCATCCG
CTCCACACCC CACCGGGTCC TGCACAACTA CCTGGTGTGG CGCGTGGTGG TGGTCCTGAG
TGAACACCTG TCCCCGCCAT TCCGTGAGGC ACTGCACGAG CTGGCACAGG AGATGGAGGG
CAGCGACAAG CCACAGGAGC TGGCCCGGGT CTGCTTGGGC CAGGCCAATC GCCACTTTGG
CATGGCGCTT GGCGCCCTCT TTGTACATGA GCACTTCTCA GCTGCCAGCA AAGCCAAGGT
GCAGCAGCTA GTGGAAGACA TCAAGTACAT CCTGGGCCAG CGCCTGGAGG AGCTGGACTG
GATGGACGCC GAGACCAGGG CTGCTGCTCG GGCCAAGCTC CAGTACATGA TGGTGATGGT
CGGCTACCCG GACTTCCTGC TGAAACCCGA TGCTGTGGAC AAGGAGTATG AGTTTGAGGT
CCATGAGAAG ACCTACTTCA AGAACATCTT GAACAGCATC CGCTTCAGCA TCCAGCTCTC

AGTTAAGAAG ATTCGGCAGG AGGTGGACAA GTCCACGTGG CTGCTCCCCC CACAGGCGCT
CAATGCCTAC TATCTACCCA ACAAGAACCA GATGGTGTTC CCCGCGGGCA TCCTGCAGCC
CACCCTGTAC GACCCTGACT TCCCACAGTC TCTCAACTAC GGGGGCATCG GCACCATCAT
TGGACATGAG CTGACCCACG GCTACGACGA CTGGGGGGGC CAGTATGACC GCTCAGGGAA
CCTGCTGCAC TGGTGGACGG AGGCCTCCTA CAGCCGCTTC CTGCGAAAGG CTGAGTGCAT
CGTCCGTCTC TATGACAACT TCACTGTCTA CAACCAGCGG GTGAACGGGA AACACACGCT
TGGGGAGAAC ATCGCAGATA TGGGCGGCCT CAAGCTGGCC TACCACGCCT ATCAGAAGTG
GGTGCGGGAG CACGGCCCAG AGCACCCACT TCCCCGGCTC AAGTACACAC ATGACCAGCT
CTTCTTCATT GCCTTTGCCC AGAACTGGTG CATCAAGCGG CGGTCGCAGT CCATCTACCT
GCAGGTGCTG ACTGACAAGC ATGCCCCTGA GCACTACAGG GTGCTGGGCA GTGTGTCCCA
GTTTGAGGAG TTTGGCCGGG CTTTCCACTG TCCCAAGGAC TCACCCATGA ACCCTGCCCA
CAAGTGTTCC GTGTGGTGAG CCTGGCTGCC CGCCTGCACG CCCCCACTGC CCCCGCACGA
ATCACCTCCT GCTGGCTACC GGGGCAGGCA TGCACCCGGT GCCAGCCCCG CTCTGGGCAC
CACCTGCCTT CCAGCCCCTC CAGGACCCGG TCCCCCTGCT GCCCCTCACT TCAGGAGGGG
CCTGGAGCAG GGTGAGGCTG GACTTTGGGG GGCTGTGAGG GAAATATACT GGGGTCCCCA
GATTCTGCTC TAAGGGGGCC AGACCCTCTG CCAGGCTGGA TTGTACGGGC CCCACCTTCG
CTGTGTTCTT GCTGCAAAGT CTGGTCAATA AATCACTGCA CTGTTAAAAA AAAAAAAAAA
AAAAAATTCC TGCG (SEQ ID NO:1).

13. The purified DNA molecule of claim 12 which consists of a nucleotide sequence from nucleotide 212 to nucleotide 2539 of SEQ ID NO:1.

14. An expression vector for the expression of endothelin converting enzyme-3 in a recombinant host cell wherein said expression vector comprises a DNA molecule of claim 13.

15. An expression vector of claim 14 which is a eukaryotic expression vector.

16. An expression vector of claim 14 which is a prokaryotic expression vector.

17. A host cell which expresses a recombinant endothelin converting enzyme-3 wherein said host cell contains the expression vector of claim 14.

18. A host cell which expresses a recombinant endothelin converting enzyme-3 wherein said host cell contains the expression vector of claim 15.

19. A host cell which expresses a recombinant endothelin converting enzyme-3 wherein said host cell contains the expression vector of claim 16.

20. A subcellular membrane fraction obtained from the host cell of claim 17 which contains recombinant endothelin converting enzyme-3.

21. A subcellular membrane fraction obtained from the host cell of claim 18 which contains recombinant endothelin converting enzyme-3.

22. A subcellular membrane fraction obtained from the host cell of claim 19 which contains recombinant endothelin converting enzyme-3.

23. A purified endothelin converting enzyme-3 protein which comprises the amino acid sequence:

Met Glu Pro Pro Tyr Ser Leu Thr Ala His Tyr Asp Glu Phe Gln Glu
Val Lys Tyr Val Ser Arg Cys Gly Ala Gly Gly Ala Arg Gly Ala Ser
Leu Pro Pro Gly Phe Pro Leu Gly Ala Ala Arg Ser Ala Thr Gly Ala
Arg Ser Gly Leu Pro Arg Trp Asn Arg Arg Glu Val Cys Leu Leu Ser
Gly Leu Val Phe Ala Ala Gly Leu Cys Ala Ile Leu Ala Ala Met Leu
Ala Leu Lys Tyr Leu Gly Pro Val Ala Ala Gly Gly Gly Ala Cys Pro
Glu Gly Cys Pro Glu Arg Lys Ala Phe Ala Arg Ala Ala Arg Phe Leu
Ala Ala Asn Leu Asp Ala Ser Ile Asp Pro Cys Gln Asp Phe Tyr Ser
Phe Ala Cys Gly Gly Trp Leu Arg Arg His Ala Ile Pro Asp Asp Lys
Leu Thr Tyr Gly Thr Ile Ala Ala Ile Gly Glu Gln Asn Glu Glu Arg
Leu Arg Arg Leu Leu Ala Arg Pro Gly Gly Gly Pro Gly Gly Ala Ala

Gln Arg Lys Val Arg Ala Phe Phe Arg Ser Cys Leu Asp Met Arg Glu
Ile Glu Arg Leu Gly Pro Arg Pro Met Leu Glu Val Ile Glu Asp Cys
Gly Gly Trp Asp Leu Gly Gly Ala Glu Glu Arg Pro Gly Val Ala Ala
Arg Trp Asp Leu Asn Arg Leu Leu Tyr Lys Ala Gln Gly Val Tyr Ser
Ala Ala Ala Leu Phe Ser Leu Thr Val Ser Leu Asp Asp Arg Asn Ser
Ser Arg Tyr Val Ile Arg Ile Asp Gln Asp Gly Leu Thr Leu Pro Glu
Arg Thr Leu Tyr Leu Ala Gln Asp Glu Asp Ser Glu Lys Ile Leu Ala
Ala Tyr Arg Val Phe Met Glu Arg Val Leu Ser Leu Leu Gly Ala Asp
Ala Val Glu Gln Lys Ala Gln Glu Ile Leu Gln Val Glu Gln Gln Leu
Ala Asn Ile Thr Val Ser Glu Tyr Asp Asp Leu Arg Arg Asp Val Ser
Ser Met Tyr Asn Lys Val Thr Leu Gly Gln Leu Gln Lys Ile Thr Pro
His Leu Arg Trp Lys Trp Leu Leu Asp Gln Ile Phe Gln Glu Asp Phe
Ser Glu Glu Glu Glu Val Val Leu Leu Ala Thr Asp Tyr Met Gln Gln
Val Ser Gln Leu Ile Arg Ser Thr Pro His Arg Val Leu His Asn Tyr
Leu Val Trp Arg Val Val Val Val Leu Ser Glu His Leu Ser Pro Pro
Phe Arg Glu Ala Leu His Glu Leu Ala Gln Glu Met Glu Gly Ser Asp
Lys Pro Gln Glu Leu Ala Arg Val Cys Leu Gly Gln Ala Asn Arg His
Phe Gly Met Ala Leu Gly Ala Leu Phe Val His Glu His Phe Ser Ala
Ala Ser Lys Ala Lys Val Gln Gln Leu Val Glu Asp Ile Lys Tyr Ile
Leu Gly Gln Arg Leu Glu Glu Leu Asp Trp Met Asp Ala Glu Thr Arg
Ala Ala Ala Arg Ala Lys Leu Gln Tyr Met Met Val Met Val Gly Tyr
Pro Asp Phe Leu Leu Lys Pro Asp Ala Val Asp Lys Glu Tyr Glu Phe
Glu Val His Glu Lys Thr Tyr Phe Lys Asn Ile Leu Asn Ser Ile Arg
Phe Ser Ile Gln Leu Ser Val Lys Lys Ile Arg Gln Glu Val Asp Lys
Ser Thr Trp Leu Leu Pro Pro Gln Ala Leu Asn Ala Tyr Tyr Leu Pro
Asn Lys Asn Gln Met Val Phe Pro Ala Gly Ile Leu Gln Pro Thr Leu
Tyr Asp Pro Asp Phe Pro Gln Ser Leu Asn Tyr Gly Gly Ile Gly Thr
Ile Ile Gly His Glu Leu Thr His Gly Tyr Asp Asp Trp Gly Gly Gln
Tyr Asp Arg Ser Gly Asn Leu Leu His Trp Trp Thr Glu Ala Ser Tyr
Ser Arg Phe Leu Arg Lys Ala Glu Cys Ile Val Arg Leu Tyr Asp Asn
Phe Thr Val Tyr Asn Gln Arg Val Asn Gly Lys His Thr Leu Gly Glu
Asn Ile Ala Asp Met Gly Gly Leu Lys Leu Ala Tyr His Ala Tyr Gln
Lys Trp Val Arg Glu His Gly Pro Glu His Pro Leu Pro Arg Leu Lys
Tyr Thr His Asp Gln Leu Phe Phe Ile Ala Phe Ala Gln Asn Trp Cys

Ile Lys Arg Arg Ser Gln Ser Ile Tyr Leu Gln Val Leu Thr Asp Lys
His Ala Pro Glu His Tyr Arg Val Leu Gly Ser Val Ser Gln Phe Glu
Glu Phe Gly Arg Ala Phe His Cys Pro Lys Asp Ser Pro Met Asn Pro
Ala His Lys Cys Ser Val Trp (SEQ ID NO:2).

24. The purified endothelin converting enzyme-3 protein of claim 23 which consists of the amino acid sequence as set forth in SEQ ID NO:2.

25. A process for the expression of an endothelin converting enzyme-3 in a recombinant host cell, comprising:

(a) transfecting the expression vector of claim 3 into a suitable host cell; and,

(b) culturing the host cells of step (a) under conditions which allow expression of the endothelin converting enzyme-3 protein from the expression vector.

26. A method for determining whether a substance is capable of binding to endothelin converting enzyme-3 comprising:

(a) providing test cells by transfecting cells with an expression vector that directs the expression of endothelin converting enzyme-3 as set forth as SEQ ID NO:2 in the cells;

(b) exposing the test cells to the substance;

(c) measuring the amount of binding of the substance to the endothelin converting enzyme-3;

(d) comparing the amount of binding of the substance to the endothelin converting enzyme-3 in the test cells with the amount of binding of the substance to control cells that have not been transfected with the expression vector that directs the expression of endothelin converting enzyme-3 as set forth as SEQ ID NO:2.

27. A method for determining whether a substance is capable of modulating endothelin converting enzyme-3 activity comprising:

(a) providing test cells by transfecting cells with an expression vector that directs the expression of endothelin converting enzyme-3 as set forth as SEQ ID NO:2;

(b) exposing the test cells to the substance;

(c) measuring the amount of an accumulated intracellular secondary message;

(d) comparing the amount of the secondary message in the test cells in response to the substance with the amount of secondary message in test cells that have not been exposed to the substance.

28. The method of claim 27 wherein the secondary message is cAMP.

29. A method for determining whether a substance is capable of modulating endothelin converting enzyme-3 activity comprising:

(a) providing test cells by transfecting cells with an expression vector that directs the expression of endothelin converting enzyme-3 as set forth as SEQ ID NO:2;

(b) purifying membrane preparations comprising the endothelin converting enzyme-3;

(c) adding a test substance to the purified membrane preparations of step (b);

(d) incubating the test substance-containing membrane preparation of step (c) with a substrate of endothelin converting enzyme-3;

(e) comparing the product generated from step (d) versus the amount of product generated from a membrane preparation containing the substrate of step (d) without addition of the test substance of step (c).

30. The method of claim 29 wherein the substrate is selected from the group consisting of big ET-1, big ET-2 and big ET-3.

GGCGGCGGGC GCTGGGAGAC ACCGGACGCC CGCTCGGCTG CGCTGCGGCT CAGGCCCCCG

CTCGGGCCCG ACCCGCTCGG TCACCGCCGG CTCGGGCGCG CACCTGCCGG CTGCGGCCCC

AGGGCCATGC GGAGGCCCAC GAGGAGGCCG GCGGCCACGC GCATCCCGTA GCCCAGGTGG

CCCAGGTCTG CACCGCGGCG GCCTCGGCGC CATGGAGCCC CCGTATTCGC TGACGGCGCA

CTACGATGAG TTCCAAGAGG TCAAGTACGT GAGCCGCTGC GGCGCGGGGG GCGCGCGCGG

GGCCTCCCTG CCCCCGGGCT TCCCGTTGGG CGCTGCCCGC AGCGCCACCG GGGCCCGGTC

CGGGCTGCCG CGCTGGAACC GGCGCGAGGT GTGCCTGCTG TCGGGGCTGG TGTTCGCCGC

CGGCCTCTGC GCCATTCTGG CGGCTATGCT GGCCCTCAAG TACCTGGGCC CGGTCGCGGC

CGGCGGCGGC GCCTGTCCCG AGGGCTGCCC TGAGCGCAAG GCCTTCGCGC GCGCCGCTCG

CTTCCTGGCC GCCAACCTGG ACGCCAGCAT CGACCCATGC CAGGACTTCT ACTCGTTCGC

CTGCGGCGGT TGGCTGCGGC GCCACGCCAT CCCCGACGAC AAGCTCACCT ATGGCACCAT

CGCGGCCATC GGCGAGCAAA ACGAGGAGCG CCTACGGCGC CTGCTGGCGC GGCCCGGGGG

TGGGCCTGGC GGCGCGGCCC AGCGCAAGGT GCGCGCCTTC TTCCGCTCGT GCCTCGACAT

GCGCGAGATC GAGCGACTGG GCCCGCGACC CATGCTAGAG GTCATCGAGG ACTGCGGGGG

CTGGGACCTG GGCGGCGCGG AGGAGCGTCC GGGGGTCGCG GCGCGATGGG ACCTCAACCG

GCTGCTGTAC AAGGCGCAGG GCGTGTACAG CGCCGCCGCG CTCTTCTCGC TCACGGTCAG

CCTGGACGAC AGGAACTCCT CGCGCTACGT CATCCGCATT GACCAGGATG GGCTCACCCT

GCCAGAGAGG ACCCTGTACC TCGCTCAGGA TGAGGACAGT GAGAAGATCC TGGCAGCATA

CAGGGTGTTC ATGGAGCGAG TGCTCAGCCT CCTGGGTGCA GACGCTGTGG AACAGAAGGC

CCAAGAGATC CTGCAAGTGG AGCAGCAGCT GGCCAACATC ACTGTGTCAG AGTATGACGA

CCTACGGCGA GATGTCAGCT CCATGTACAA CAAGGTGACG CTGGGGCAGC TGCAGAAGAT

CACCCCCCAC TTGCGGTGGA AGTGGCTGCT AGACCAGATC TTCCAGGAGG ACTTCTCAGA

GGAAGAGGAG GTGGTGCTGC TGGCGACAGA CTACATGCAG CAGGTGTCGC AGCTCATCCG

CTCCACACCC CACCGGGTCC TGCACAACTA CCTGGTGTGG CGCGTGGTGG TGGTCCTGAG

TGAACACCTG TCCCCGCCAT TCCGTGAGGC ACTGCACGAG CTGGCACAGG AGATGGAGGG

CAGCGACAAG CCACAGGAGC TGGCCCGGGT CTGCTTGGGC CAGGCCAATC GCCACTTTGG

FIG. 1A

CATGGCGCTT GGCGCCCTCT TTGTACATGA GCACTTCTCA GCTGCCAGCA AAGCCAAGGT
GCAGCAGCTA GTGGAAGACA TCAAGTACAT CCTGGGCCAG CGCCTGGAGG AGCTGGACTG
GATGGACGCC GAGACCAGGG CTGCTGCTCG GGCCAAGCTC CAGTACATGA TGGTGATGGT
CGGCTACCCG GACTTCCTGC TGAAACCCGA TGCTGTGGAC AAGGAGTATG AGTTTGAGGT
CCATGAGAAG ACCTACTTCA AGAACATCTT GAACAGCATC CGCTTCAGCA TCCAGCTCTC
AGTTAAGAAG ATTCGGCAGG AGGTGGACAA GTCCACGTGG CTGCTCCCCC CACAGGCGCT
CAATGCCTAC TATCTACCCA ACAAGAACCA GATGGTGTTC CCCGCGGGCA TCCTGCAGCC
CACCCTGTAC GACCCTGACT TCCCACAGTC TCTCAACTAC GGGGGCATCG GCACCATCAT
TGGACATGAG CTGACCCACG GCTACGACGA CTGGGGGGGC CAGTATGACC GCTCAGGGAA
CCTGCTGCAC TGGTGGACGG AGGCCTCCTA CAGCCGCTTC CTGCGAAAGG CTGAGTGCAT
CGTCCGTCTC TATGACAACT TCACTGTCTA CAACCAGCGG GTGAACGGGA AACACACGCT
TGGGGAGAAC ATCGCAGATA TGGGCGGCCT CAAGCTGGCC TACCACGCCT ATCAGAAGTG
GGTGCGGGAG CACGGCCCAG AGCACCCACT TCCCCGGCTC AAGTACACAC ATGACCAGCT
CTTCTTCATT GCCTTTGCCC AGAACTGGTG CATCAAGCGG CGGTCGCAGT CCATCTACCT
GCAGGTGCTG ACTGACAAGC ATGCCCCTGA GCACTACAGG GTGCTGGGCA GTGTGTCCCA
GTTTGAGGAG TTTGGCCGGG CTTTCCACTG TCCCAAGGAC TCACCCATGA ACCCTGCCCA
CAAGTGTTCC GTGTGGTGAG CCTGGCTGCC CGCCTGCACG CCCCCACTGC CCCCGCACGA
ATCACCTCCT GCTGGCTACC GGGGCAGGCA TGCACCCGGT GCCAGCCCCG CTCTGGGCAC
CACCTGCCTT CCAGCCCCTC CAGGACCCGG TCCCCCTGCT GCCCCTCACT TCAGGAGGGG
CCTGGAGCAG GGTGAGGCTG GACTTTGGGG GGCTGTGAGG GAAATATACT GGGGTCCCCA
GATTCTGCTC TAAGGGGGCC AGACCCTCTG CCAGGCTGGA TTGTACGGGC CCCACCTTCG
CTGTGTTCTT GCTGCAAAGT CTGGTCAATA AATCACTGCA CTGTTAAAAA AAAAAAAAAA
AAAAAATTCC TGCG (SEQ ID NO:1)

FIG.1B

Met Glu Pro Pro Tyr Ser Leu Thr Ala His Tyr Asp Glu Phe Gln Glu
Val Lys Tyr Val Ser Arg Cys Gly Ala Gly Gly Ala Arg Gly Ala Ser
Leu Pro Pro Gly Phe Pro Leu Gly Ala Ala Arg Ser Ala Thr Gly Ala
Arg Ser Gly Leu Pro Arg Trp Asn Arg Arg Glu Val Cys Leu Leu Ser
Gly Leu Val Phe Ala Ala Gly Leu Cys Ala Ile Leu Ala Ala Met Leu
Ala Leu Lys Tyr Leu Gly Pro Val Ala Ala Gly Gly Gly Ala Cys Pro
Glu Gly Cys Pro Glu Arg Lys Ala Phe Ala Arg Ala Ala Arg Phe Leu
Ala Ala Asn Leu Asp Ala Ser Ile Asp Pro Cys Gln Asp Phe Tyr Ser
Phe Ala Cys Gly Gly Trp Leu Arg Arg His Ala Ile Pro Asp Asp Lys
Leu Thr Tyr Gly Thr Ile Ala Ala Ile Gly Glu Gln Asn Glu Glu Arg
Leu Arg Arg Leu Leu Ala Arg Pro Gly Gly Gly Pro Gly Gly Ala Ala
Gln Arg Lys Val Arg Ala Phe Phe Arg Ser Cys Leu Asp Met Arg Glu
Ile Glu Arg Leu Gly Pro Arg Pro Met Leu Glu Val Ile Glu Asp Cys
Gly Gly Trp Asp Leu Gly Gly Ala Glu Glu Arg Pro Gly Val Ala Ala
Arg Trp Asp Leu Asn Arg Leu Leu Tyr Lys Ala Gln Gly Val Tyr Ser
Ala Ala Ala Leu Phe Ser Leu Thr Val Ser Leu Asp Asp Arg Asn Ser
Ser Arg Tyr Val Ile Arg Ile Asp Gln Asp Gly Leu Thr Leu Pro Glu
Arg Thr Leu Tyr Leu Ala Gln Asp Glu Asp Ser Glu Lys Ile Leu Ala
Ala Tyr Arg Val Phe Met Glu Arg Val Leu Ser Leu Leu Gly Ala Asp
Ala Val Glu Gln Lys Ala Gln Glu Ile Leu Gln Val Glu Gln Gln Leu
Ala Asn Ile Thr Val Ser Glu Tyr Asp Asp Leu Arg Arg Asp Val Ser
Ser Met Tyr Asn Lys Val Thr Leu Gly Gln Leu Gln Lys Ile Thr Pro
His Leu Arg Trp Lys Trp Leu Leu Asp Gln Ile Phe Gln Glu Asp Phe
Ser Glu Glu Glu Glu Val Val Leu Leu Ala Thr Asp Tyr Met Gln Gln
Val Ser Gln Leu Ile Arg Ser Thr Pro His Arg Val Leu His Asn Tyr
Leu Val Trp Arg Val Val Val Val Leu Ser Glu His Leu Ser Pro Pro

FIG.2A

Phe Arg Glu Ala Leu His Glu Leu Ala Gln Glu Met Glu Gly Ser Asp
Lys Pro Gln Glu Leu Ala Arg Val Cys Leu Gly Gln Ala Asn Arg His
Phe Gly Met Ala Leu Gly Ala Leu Phe Val His Glu His Phe Ser Ala
Ala Ser Lys Ala Lys Val Gln Gln Leu Val Glu Asp Ile Lys Tyr Ile
Leu Gly Gln Arg Leu Glu Glu Leu Asp Trp Met Asp Ala Glu Thr Arg
Ala Ala Ala Arg Ala Lys Leu Gln Tyr Met Met Val Met Val Gly Tyr
Pro Asp Phe Leu Leu Lys Pro Asp Ala Val Asp Lys Glu Tyr Glu Phe
Glu Val His Glu Lys Thr Tyr Phe Lys Asn Ile Leu Asn Ser Ile Arg
Phe Ser Ile Gln Leu Ser Val Lys Lys Ile Arg Gln Glu Val Asp Lys
Ser Thr Trp Leu Leu Pro Pro Gln Ala Leu Asn Ala Tyr Tyr Leu Pro
Asn Lys Asn Gln Met Val Phe Pro Ala Gly Ile Leu Gln Pro Thr Leu
Tyr Asp Pro Asp Phe Pro Gln Ser Leu Asn Tyr Gly Gly Ile Gly Thr
Ile Ile Gly His Glu Leu Thr His Gly Tyr Asp Asp Trp Gly Gly Gln
Tyr Asp Arg Ser Gly Asn Leu Leu His Trp Trp Thr Glu Ala Ser Tyr
Ser Arg Phe Leu Arg Lys Ala Glu Cys Ile Val Arg Leu Tyr Asp Asn
Phe Thr Val Tyr Asn Gln Arg Val Asn Gly Lys His Thr Leu Gly Glu
Asn Ile Ala Asp Met Gly Gly Leu Lys Leu Ala Tyr His Ala Tyr Gln
Lys Trp Val Arg Glu His Gly Pro Glu His Pro Leu Pro Arg Leu Lys
Tyr Thr His Asp Gln Leu Phe Phe Ile Ala Phe Ala Gln Asn Trp Cys
Ile Lys Arg Arg Ser Gln Ser Ile Tyr Leu Gln Val Leu Thr Asp Lys
His Ala Pro Glu His Tyr Arg Val Leu Gly Ser Val Ser Gln Phe Glu
Glu Phe Gly Arg Ala Phe His Cys Pro Lys Asp Ser Pro Met Asn Pro
Ala His Lys Cys Ser Val Trp (SEQ ID NO:2)

FIG.2B

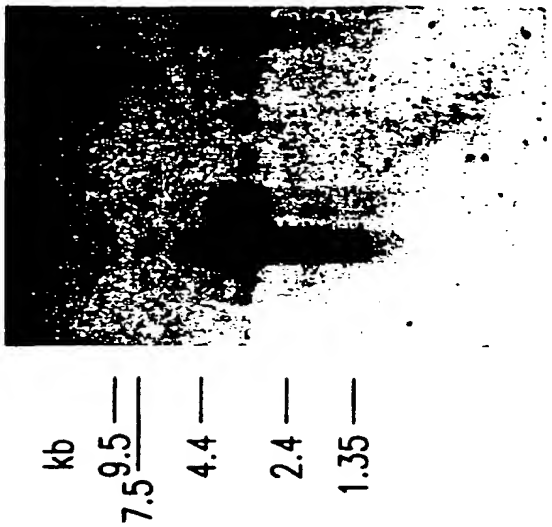
kb
9.5
7.5
4.4
2.4
1.35
cerebellum
cerebral cortex
medulla
spinal cord
occipital pole
frontal lobe
temporal lobe
putamen

FIG.3B


kb
9.5
7.5
4.4
2.4
1.35
amygdala
caudate nucleus
corpus callosum
hippocampus
whole brain
substantia nigra
subthalamic nucleus
thalamus

FIG.3A

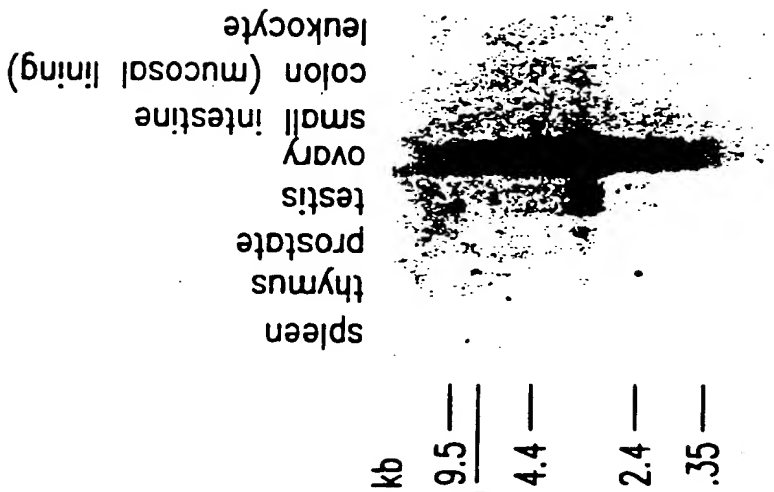
kb
9.5
7.5
4.4
2.4
1.35
spleen
thymus
prostate
testis
ovary
small intestine
colon (mucosal lining)
leukocyte

FIG.3D

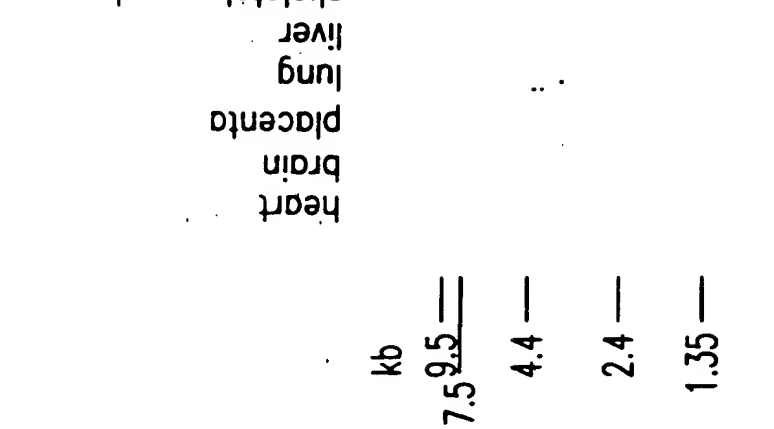
kb
9.5
7.5
4.4
2.4
1.35
heart
brain
placenta
lung
liver
skeletal muscle
kidney
pancreas

FIG.3C

```
          10         20         30         40         50
GGCGGCGGGCGCTGGGAGACACCGGACGCCCGCTCGGCTGCGCTGCGGCT
CCGCCGCCCGCGACCCTCTGTGGCCTGCGGGCGAGCCGACGCGACGCCGA

          60         70         80         90        100
CAGGCCCCCGCTCGGGCCCGACCCGCTCGGTCACCGCCGGCTCGGGCGCG
GTCCGGGGGCGAGCCCGGGCTGGGCGAGCCAGTGGCGGCCGAGCCCGCGC

         110        120        130        140        150
CACCTGCCGGCTGCGGCCCCAGGGCCATGCGGAGGCCCACGAGGAGGCCG
GTGGACGGCCGACGCCGGGGTCCCGGTACGCCTCCGGGTGCTCCTCCGGC

         160        170        180        190        200
GCGGCCACGCGCATCCCGTAGCCCAGGTGGCCCAGGTCTGCACCGCGGCG
CGCCGGTGCGCGTAGGGCATCGGGTCCACCGGGTCCAGACGTGGCGCCGC

         210        220        230        240        250
GCCTCGGCGCCATGGAGCCCCCGTATTCGCTGACGGCGCACTACGATGAG
CGGAGCCGCGGTACCTCGGGGGCATAAGCGACTGCCGCGTGATGCTACTC
               M  E  P  P  Y  S  L  T  A  H  Y  D  E>

         260        270        280        290        300
TTCCAAGAGGTCAAGTACGTGAGCCGCTGCGGCGCGGGGGGCGCGCGCGG
AAGGTTCTCCAGTTCATGCACTCGGCGACGCCGCGCCCCCCGCGCGCGCC
 F  Q  E  V  K  Y  V  S  R  C  G  A  G  G  A  R  G>

         310        320        330        340        350
GGCCTCCCTGCCCCCGGGCTTCCCGTTGGGCGCTGCCCGCAGCGCCACCG
CCGGAGGGACGGGGGCCCGAAGGGCAACCCGCGACGGGCGTCGCGGTGGC
  A  S  L  P  P  G  F  P  L  G  A  A  R  S  A  T>

         360        370        380        390        400
GGGCCCGGTCCGGGCTGCCGCGCTGGAACCGGCGCGAGGTGTGCCTGCTG
CCCGGGCCAGGCCCGACGGCGCGACCTTGGCCGCGCTCCACACGGACGAC
G  A  R  S  G  L  P  R  W  N  R  R  E  V  C  L  L>

         410        420        430        440        450
TCGGGGCTGGTGTTCGCCGCCGGCCTCTGCGCCATTCTGGCGGCTATGCT
AGCCCCGACCACAAGCGGCGGCCGGAGACGCGGTAAGACCGCCGATACGA
 S  G  L  V  F  A  A  G  L  C  A  I  L  A  A  M  L>
```

FIG.4A

```
        460        470        480        490        500
GGCCCTCAAGTACCTGGGCCCGGTCGCGGCCGGCGGCGGCGCCTGTCCCG
CCGGGAGTTCATGGACCCGGGCCAGCGCCGGCCGCCGCCGCGGACAGGGC
  A  L  K  Y  L  G  P  V  A  A  G  G  G  A  C  P>

        510        520        530        540        550
AGGGCTGCCCTGAGCGCAAGGCCTTCGCGCGCCGCTCGCTTCCTGGCC
TCCCGACGGGACTCGCGTTCCGGAAGCGCGCGCGGCGAGCGAAGGACCGG
E  G  C  P  E  R  K  A  F  A  R  A  A  R  F  L  A>

        560        570        580        590        600
GCCAACCTGGACGCCAGCATCGACCCATGCCAGGACTTCTACTCGTTCGC
CGGTTGGACCTGCGGTCGTAGCTGGGTACGGTCCTGAAGATGAGCAAGCG
 A  N  L  D  A  S  I  D  P  C  Q  D  F  Y  S  F  A>

        610        620        630        640        650
CTGCGGCGGTTGGCTGCGGCGCCACGCCATCCCCGACGACAAGCTCACCT
GACGCCGCCAACCGACGCCGCGGTGCGGTAGGGGCTGCTGTTCGAGTGGA
  C  G  G  W  L  R  R  H  A  I  P  D  D  K  L  T>

        660        670        680        690        700
ATGGCACCATCGCGGCCATCGGCGAGCAAAACGAGGAGCGCCTACGGCGC
TACCGTGGTAGCGCCGGTAGCCGCTCGTTTTGCTCCTCGCGGATGCCGCG
Y  G  T  I  A  A  I  G  E  Q  N  E  E  R  L  R  R>

        710        720        730        740        750
CTGCTGGCGCGGCCCGGGGGTGGGCCTGGCGGCGCGGCCCAGCGCAAGGT
GACGACCGCGCCGGGCCCCCACCCGGACCGCCGCGCCGGGTCGCGTTCCA
 L  L  A  R  P  G  G  G  P  G  G  A  A  Q  R  K  V>

        760        770        780        790        800
GCGCGCCTTCTTCCGCTCGTGCCTCGACATGCGCGAGATCGAGCGACTGG
CGCGCGGAAGAAGGCGAGCACGGAGCTGTACGCGCTCTAGCTCGCTGACC
  R  A  F  F  R  S  C  L  D  M  R  E  I  E  R  L>

        810        820        830        840        850
GCCCGCGACCCATGCTAGAGGTCATCGAGGACTGCGGGGGCTGGGACCTG
CGGGCGCTGGGTACGATCTCCAGTAGCTCCTGACGCCCCCGACCCTGGAC
G  P  R  P  M  L  E  V  I  E  D  C  G  G  W  D  L>
```

FIG.4B

```
        860        870        880        890        900
GGCGGCGCGGAGGAGCGTCCGGGGGTCGCGGCGCGATGGGACCTCAACCG
CCGCCGCGCCTCCTCGCAGGCCCCCAGCGCCGCGCTACCCTGGAGTTGGC
 G  G  A  E  E  R  P  G  V  A  A  R  W  D  L  N  R>

        910        920        930        940        950
GCTGCTGTACAAGGCGCAGGGCGTGTACAGCGCCGCCGCGCTCTTCTCGC
CGACGACATGTTCCGCGTCCCGCACATGTCGCGGCGGCGCGAGAAGAGCG
  L  L  Y  K  A  Q  G  V  Y  S  A  A  A  L  F  S>

        960        970        980        990       1000
TCACGGTCAGCCTGGACGACAGGAACTCCTCGCGCTACGTCATCCGCATT
AGTGCCAGTCGGACCTGCTGTCCTTGAGGAGCGCGATGCAGTAGGCGTAA
L  T  V  S  L  D  D  R  N  S  S  R  Y  V  I  R  I>

       1010       1020       1030       1040       1050
GACCAGGATGGGCTCACCCTGCCAGAGAGGACCCTGTACCTCGCTCAGGA
CTGGTCCTACCCGAGTGGGACGGTCTCTCCTGGGACATGGAGCGAGTCCT
 D  Q  D  G  L  T  L  P  E  R  T  L  Y  L  A  Q  D>

       1060       1070       1080       1090       1100
TGAGGACAGTGAGAAGATCCTGGCAGCATACAGGGTGTTCATGGAGCGAG
ACTCCTGTCACTCTTCTAGGACCGTCGTATGTCCCACAAGTACCTCGCTC
  E  D  S  E  K  I  L  A  A  Y  R  V  F  M  E  R>

       1110       1120       1130       1140       1150
TGCTCAGCCTCCTGGGTGCAGACGCTGTGGAACAGAAGGCCCAAGAGATC
ACGAGTCGGAGGACCCACGTCTGCGACACCTTGTCTTCCGGGTTCTCTAG
V  L  S  L  L  G  A  D  A  V  E  Q  K  A  Q  E  I>

       1160       1170       1180       1190       1200
CTGCAAGTGGAGCAGCAGCTGGCCAACATCACTGTGTCAGAGTATGACGA
GACGTTCACCTCGTCGTCGACCGGTTGTAGTGACACAGTCTCATACTGCT
 L  Q  V  E  Q  Q  L  A  N  I  T  V  S  E  Y  D  D>

       1210       1220       1230       1240       1250
CCTACGGCGAGATGTCAGCTCCATGTACAACAAGGTGACGCTGGGGCAGC
GGATGCCGCTCTACAGTCGAGGTACATGTTGTTCCACTGCGACCCCGTCG
  L  R  R  D  V  S  S  M  Y  N  K  V  T  L  G  Q>
```

FIG.4C

```
      1260       1270       1280       1290       1300
TGCAGAAGATCACCCCCCACTTGCGGTGGAAGTGGCTGCTAGACCAGATC
ACGTCTTCTAGTGGGGGGTGAACGCCACCTTCACCGACGATCTGGTCTAG
L  Q  K  I  T  P  H  L  R  W  K  W  L  L  D  Q  I>

      1310       1320       1330       1340       1350
TTCCAGGAGGACTTCTCAGAGGAAGAGGAGGTGGTGCTGCTGGCGACAGA
AAGGTCCTCCTGAAGAGTCTCCTTCTCCTCCACCACGACGACCGCTGTCT
 F  Q  E  D  F  S  E  E  E  E  V  V  L  L  A  T  D>

      1360       1370       1380       1390       1400
CTACATGCAGCAGGTGTCGCAGCTCATCCGCTCCACACCCCACCGGGTCC
GATGTACGTCGTCCACAGCGTCGAGTAGGCGAGGTGTGGGGTGGCCCAGG
  Y  M  Q  Q  V  S  Q  L  I  R  S  T  P  H  R  V>

      1410       1420       1430       1440       1450
TGCACAACTACCTGGTGTGGCGCGTGGTGGTGGTCCTGAGTGAACACCTG
ACGTGTTGATGGACCACACCGCGCACCACCACCAGGACTCACTTGTGGAC
L  H  N  Y  L  V  W  R  V  V  V  V  L  S  E  H  L>

      1460       1470       1480       1490       1500
TCCCCGCCATTCCGTGAGGCACTGCACGAGCTGGCACAGGAGATGGAGGG
AGGGGCGGTAAGGCACTCCGTGACGTGCTCGACCGTGTCCTCTACCTCCC
 S  P  P  F  R  E  A  L  H  E  L  A  Q  E  M  E  G>

      1510       1520       1530       1540       1550
CAGCGACAAGCCACAGGAGCTGGCCCGGGTCTGCTTGGGCCAGGCCAATC
GTCGCTGTTCGGTGTCCTCGACCGGGCCCAGACGAACCCGGTCCGGTTAG
  S  D  K  P  Q  E  L  A  R  V  C  L  G  Q  A  N>

      1560       1570       1580       1590       1600
GCCACTTTGGCATGGCGCTTGGCGCCCTCTTTGTACATGAGCACTTCTCA
CGGTGAAACCGTACCGCGAACCGCGGGAGAAACATGTACTCGTGAAGAGT
R  H  F  G  M  A  L  G  A  L  F  V  H  E  H  F  S>

      1610       1620       1630       1640       1650
GCTGCCAGCAAAGCCAAGGTGCAGCAGCTAGTGGAAGACATCAAGTACAT
CGACGGTCGTTTCGGTTCCACGTCGTCGATCACCTTCTGTAGTTCATGTA
 A  A  S  K  A  K  V  Q  Q  L  V  E  D  I  K  Y  I>
```

FIG.4D

```
      1660       1670       1680       1690       1700
CCTGGGCCAGCGCCTGGAGGAGCTGGACTGGATGGACGCCGAGACCAGGG
GGACCCGGTCGCGGACCTCCTCGACCTGACCTACCTGCGGCTCTGGTCCC
  L  G  Q  R  L  E  E  L  D  W  M  D  A  E  T  R>

      1710       1720       1730       1740       1750
CTGCTGCTCGGGCCAAGCTCCAGTACATGATGGTGATGGTCGGCTACCCG
GACGACGAGCCCGGTTCGAGGTCATGTACTACCACTACCAGCCGATGGGC
A  A  A  R  A  K  L  Q  Y  M  M  V  M  V  G  Y  P>

      1760       1770       1780       1790       1800
GACTTCCTGCTGAAACCCGATGCTGTGGACAAGGAGTATGAGTTTGAGGT
CTGAAGGACGACTTTGGGCTACGACACCTGTTCCTCATACTCAAACTCCA
 D  F  L  L  K  P  D  A  V  D  K  E  Y  E  F  E  V>

      1810       1820       1830       1840       1850
CCATGAGAAGACCTACTTCAAGAACATCTTGAACAGCATCCGCTTCAGCA
GGTACTCTTCTGGATGAAGTTCTTGTAGAACTTGTCGTAGGCGAAGTCGT
  H  E  K  T  Y  F  K  N  I  L  N  S  I  R  F  S>

      1860       1870       1880       1890       1900
TCCAGCTCTCAGTTAAGAAGATTCGGCAGGAGGTGGACAAGTCCACGTGG
AGGTCGAGAGTCAATTCTTCTAAGCCGTCCTCCACCTGTTCAGGTGCACC
I  Q  L  S  V  K  K  I  R  Q  E  V  D  K  S  T  W>

      1910       1920       1930       1940       1950
CTGCTCCCCCCACAGGCGCTCAATGCCTACTATCTACCCAACAAGAACCA
GACGAGGGGGGGTGTCCGCGAGTTACGGATGATAGATGGGTTGTTCTTGGT
 L  L  P  P  Q  A  L  N  A  Y  Y  L  P  N  K  N  Q>

      1960       1970       1980       1990       2000
GATGGTGTTCCCCGCGGGCATCCTGCAGCCCACCCTGTACGACCCTGACT
CTACCACAAGGGGCGCCCGTAGGACGTCGGGTGGGACATGCTGGGACTGA
  M  V  F  P  A  G  I  L  Q  P  T  L  Y  D  P  D>

      2010       2020       2030       2040       2050
TCCCACAGTCTCTCAACTACGGGGGCATCGGCACCATCATTGGACATGAG
AGGGTGTCAGAGAGTTGATGCCCCCGTAGCCGTGGTAGTAACCTGTACTC
F  P  Q  S  L  N  Y  G  G  I  G  T  I  I  G  H  E>
```

FIG.4E

```
      2060       2070       2080       2090       2100
CTGACCCACGGCTACGACGACTGGGGGGGCCAGTATGACCGCTCAGGGAA
GACTGGGTGCCGATGCTGCTGACCCCCCCGGTCATACTGGCGAGTCCCTT
 L  T  H  G  Y  D  D  W  G  G  Q  Y  D  R  S  G  N>

      2110       2120       2130       2140       2150
CCTGCTGCACTGGTGGACGGAGGCCTCCTACAGCCGCTTCCTGCGAAAGG
GGACGACGTGACCACCTGCCTCCGGAGGATGTCGGCGAAGGACGCTTTCC
 L  L  H  W  W  T  E  A  S  Y  S  R  F  L  R  K>

      2160       2170       2180       2190       2200
CTGAGTGCATCGTCCGTCTCTATGACAACTTCACTGTCTACAACCAGCGG
GACTCACGTAGCAGGCAGAGATACTGTTGAAGTGACAGATGTTGGTCGCC
A  E  C  I  V  R  L  Y  D  N  F  T  V  Y  N  Q  R>

      2210       2220       2230       2240       2250
GTGAACGGGAAACACACGCTTGGGGAGAACATCGCAGATATGGGCGGCCT
CACTTGCCCTTTGTGTGCGAACCCCTCTTGTAGCGTCTATACCCGCCGGA
 V  N  G  K  H  T  L  G  E  N  I  A  D  M  G  G  L>

      2260       2270       2280       2290       2300
CAAGCTGGCCTACCACGCCTATCAGAAGTGGGTGCGGGAGCACGGCCCAG
GTTCGACCGGATGGTGCGGATAGTCTTCACCCACGCCCTCGTGCCGGGTC
 K  L  A  Y  H  A  Y  Q  K  W  V  R  E  H  G  P>

      2310       2320       2330       2340       2350
AGCACCCACTTCCCCGGCTCAAGTACACACATGACCAGCTCTTCTTCATT
TCGTGGGTGAAGGGGCCGAGTTCATGTGTGTACTGGTCGAGAAGAAGTAA
E  H  P  L  P  R  L  K  Y  T  H  D  Q  L  F  F  I>

      2360       2370       2380       2390       2400
GCCTTTGCCCAGAACTGGTGCATCAAGCGGCGGTCGCAGTCCATCTACCT
CGGAAACGGGTCTTGACCACGTAGTTCGCCGCCAGCGTCAGGTAGATGGA
 A  F  A  Q  N  W  C  I  K  R  R  S  Q  S  I  Y  L>

      2410       2420       2430       2440       2450
GCAGGTGCTGACTGACAAGCATGCCCCTGAGCACTACAGGGTGCTGGGCA
CGTCCACGACTGACTGTTCGTACGGGGACTCGTGATGTCCCACGACCCGT
 Q  V  L  T  D  K  H  A  P  E  H  Y  R  V  L  G>
```

FIG.4F

```
      2460       2470       2480       2490       2500
GTGTGTCCCAGTTTGAGGAGTTTGGCCGGGCTTTCCACTGTCCCAAGGAC
CACACAGGGTCAAACTCCTCAAACCGGCCCGAAAGGTGACAGGGTTCCTG
S  V  S  Q  F  E  E  F  G  R  A  F  H  C  P  K  D>

      2510       2520       2530       2540       2550
TCACCCATGAACCCTGCCCACAAGTGTTCCGTGTGGTGAGCCTGGCTGCC
AGTGGGTACTTGGGACGGGTGTTCACAAGGCACACCACTCGGACCGACGG
 S  P  M  N  P  A  H  K  C  S  V  W  * (SEQ ID NO:2)

      2560       2570       2580       2590       2600
CGCCTGCACGCCCCCACTGCCCCCGCACGAATCACCTCCTGCTGGCTACC
GCGGACGTGCGGGGGTGACGGGGGCGTGCTTAGTGGAGGACGACCGATGG

      2610       2620       2630       2640       2650
GGGGCAGGCATGCACCCGGTGCCAGCCCCGCTCTGGGCACCACCTGCCTT
CCCCGTCCGTACGTGGGCCACGGTCGGGGCGAGACCCGTGGTGGACGGAA

      2660       2670       2680       2690       2700
CCAGCCCCTCCAGGACCCGGTCCCCCTGCTGCCCCTCACTTCAGGAGGGG
GGTCGGGGAGGTCCTGGGCCAGGGGGACGACGGGGAGTGAAGTCCTCCCC

      2710       2720       2730       2740       2750
CCTGGAGCAGGGTGAGGCTGGACTTTGGGGGGCTGTGAGGGAAATATACT
GGACCTCGTCCCACTCCGACCTGAAACCCCCCGACACTCCCTTTATATGA

      2760       2770       2780       2790       2800
GGGGTCCCCAGATTCTGCTCTAAGGGGGCCAGACCCTCTGCCAGGCTGGA
CCCCAGGGGTCTAAGACGAGATTCCCCCGGTCTGGGAGACGGTCCGACCT

      2810       2820       2830       2840       2850
TTGTACGGGCCCCACCTTCGCTGTGTTCTTGCTGCAAAGTCTGGTCAATA
AACATGCCCGGGGTGGAAGCGACACAAGAACGACGTTTCAGACCAGTTAT

      2860       2870       2880       2890
AATCACTGCACTGTTAAAAAAAAAAAAAAAAAAAAATTCCTGCG (SEQ ID NO:1)
TTAGTGACGTGACAATTTTTTTTTTTTTTTTTTTTTAAGGACGC (SEQ ID NO:22)
```

FIG.4G

SEQUENCE LISTING

<110> APPLICANT: Merck & Co., Inc.

<120> TITLE: DNA MOLECULES ENCODING HUMAN ENDOTHELIN CONVERTING ENZYME 3

<130> DOCKET/FILE REFERENCE: 20423

<160> NUMBER OF SEQUENCES: 22

<170> SOFTWARE: FastSEQ for Windows Version 4.0

<210> SEQ ID NO:1
<211> LENGTH: 2894
<212> TYPE: DNA
<213> ORGANISM:Homo sapien

<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (212)...(2536)

<400> SEQ ID NO:1

```
ggcggcgggc gctgggagac accggacgcc cgctcggctg cgctgcggct caggcccccg      60
ctcgggcccg acccgctcgg tcaccgccgg ctcgggcgcg cacctgccgg ctgcggcccc     120
agggccatgc ggaggcccac gaggaggccg gcggccacgc gcatcccgta gcccaggtgg     180
cccaggtctg caccgcggcg gcctcggcgc c atg gag ccc ccg tat tcg ctg        232
                                    Met Glu Pro Pro Tyr Ser Leu
                                     1               5

acg gcg cac tac gat gag ttc caa gag gtc aag tac gtg agc cgc tgc      280
Thr Ala His Tyr Asp Glu Phe Gln Glu Val Lys Tyr Val Ser Arg Cys
         10                  15                  20

ggc gcg ggg ggc gcg cgc ggg gcc tcc ctg ccc ccg ggc ttc ccg ttg      328
Gly Ala Gly Gly Ala Arg Gly Ala Ser Leu Pro Pro Gly Phe Pro Leu
     25                  30                  35

ggc gct gcc cgc agc gcc acc ggg gcc cgg tcc ggg ctg ccg cgc tgg      376
Gly Ala Ala Arg Ser Ala Thr Gly Ala Arg Ser Gly Leu Pro Arg Trp
 40                  45                  50                  55

aac cgg cgc gag gtg tgc ctg ctg tcg ggg ctg gtg ttc gcc gcc ggc      424
Asn Arg Arg Glu Val Cys Leu Leu Ser Gly Leu Val Phe Ala Ala Gly
                 60                  65                  70

ctc tgc gcc att ctg gcg gct atg ctg gcc ctc aag tac ctg ggc ccg      472
Leu Cys Ala Ile Leu Ala Ala Met Leu Ala Leu Lys Tyr Leu Gly Pro
             75                  80                  85

gtc gcg gcc ggc ggc ggc gcc tgt ccc gag ggc tgc cct gag cgc aag      520
Val Ala Ala Gly Gly Gly Ala Cys Pro Glu Gly Cys Pro Glu Arg Lys
         90                  95                  100

gcc ttc gcg cgc gcc gct cgc ttc ctg gcc gcc aac ctg gac gcc agc      568
Ala Phe Ala Arg Ala Ala Arg Phe Leu Ala Ala Asn Leu Asp Ala Ser
    105                 110                 115
```

```
atc gac cca tgc cag gac ttc tac tcg ttc gcc tgc ggc ggt tgg ctg      616
Ile Asp Pro Cys Gln Asp Phe Tyr Ser Phe Ala Cys Gly Gly Trp Leu
120                 125                 130                 135

cgg cgc cac gcc atc ccc gac gac aag ctc acc tat ggc acc atc gcg      664
Arg Arg His Ala Ile Pro Asp Asp Lys Leu Thr Tyr Gly Thr Ile Ala
                140                 145                 150

gcc atc ggc gag caa aac gag gag cgc cta cgg cgc ctg ctg gcg cgg      712
Ala Ile Gly Glu Gln Asn Glu Glu Arg Leu Arg Arg Leu Leu Ala Arg
            155                 160                 165

ccc ggg ggt ggg cct ggc ggc gcg gcc cag cgc aag gtg cgc gcc ttc      760
Pro Gly Gly Gly Pro Gly Gly Ala Ala Gln Arg Lys Val Arg Ala Phe
        170                 175                 180

ttc cgc tcg tgc ctc gac atg cgc gag atc gag cga ctg ggc ccg cga      808
Phe Arg Ser Cys Leu Asp Met Arg Glu Ile Glu Arg Leu Gly Pro Arg
    185                 190                 195

ccc atg cta gag gtc atc gag gac tgc ggg ggc tgg gac ctg ggc ggc      856
Pro Met Leu Glu Val Ile Glu Asp Cys Gly Gly Trp Asp Leu Gly Gly
200                 205                 210                 215

gcg gag gag cgt ccg ggg gtc gcg gcg cga tgg gac ctc aac cgg ctg      904
Ala Glu Glu Arg Pro Gly Val Ala Ala Arg Trp Asp Leu Asn Arg Leu
                220                 225                 230

ctg tac aag gcg cag ggc gtg tac agc gcc gcc gcg ctc ttc tcg ctc      952
Leu Tyr Lys Ala Gln Gly Val Tyr Ser Ala Ala Ala Leu Phe Ser Leu
            235                 240                 245

acg gtc agc ctg gac gac agg aac tcc tcg cgc tac gtc atc cgc att     1000
Thr Val Ser Leu Asp Asp Arg Asn Ser Ser Arg Tyr Val Ile Arg Ile
        250                 255                 260

gac cag gat ggg ctc acc ctg cca gag agg acc ctg tac ctc gct cag     1048
Asp Gln Asp Gly Leu Thr Leu Pro Glu Arg Thr Leu Tyr Leu Ala Gln
    265                 270                 275

gat gag gac agt gag aag atc ctg gca gca tac agg gtg ttc atg gag     1096
Asp Glu Asp Ser Glu Lys Ile Leu Ala Ala Tyr Arg Val Phe Met Glu
280                 285                 290                 295

cga gtg ctc agc ctc ctg ggt gca gac gct gtg gaa cag aag gcc caa     1144
Arg Val Leu Ser Leu Leu Gly Ala Asp Ala Val Glu Gln Lys Ala Gln
                300                 305                 310

gag atc ctg caa gtg gag cag cag ctg gcc aac atc act gtg tca gag     1192
Glu Ile Leu Gln Val Glu Gln Gln Leu Ala Asn Ile Thr Val Ser Glu
            315                 320                 325

tat gac gac cta cgg cga gat gtc agc tcc atg tac aac aag gtg acg     1240
Tyr Asp Asp Leu Arg Arg Asp Val Ser Ser Met Tyr Asn Lys Val Thr
        330                 335                 340

ctg ggg cag ctg cag aag atc acc ccc cac ttg cgg tgg aag tgg ctg     1288
Leu Gly Gln Leu Gln Lys Ile Thr Pro His Leu Arg Trp Lys Trp Leu
    345                 350                 355
```

```
cta gac cag atc ttc cag gag gac ttc tca gag gaa gag gag gtg gtg      1336
Leu Asp Gln Ile Phe Gln Glu Asp Phe Ser Glu Glu Glu Glu Val Val
360                 365                 370                 375

ctg ctg gcg aca gac tac atg cag cag gtg tcg cag ctc atc cgc tcc      1384
Leu Leu Ala Thr Asp Tyr Met Gln Gln Val Ser Gln Leu Ile Arg Ser
                380                 385                 390

aca ccc cac cgg gtc ctg cac aac tac ctg gtg tgg cgc gtg gtg gtg      1432
Thr Pro His Arg Val Leu His Asn Tyr Leu Val Trp Arg Val Val Val
            395                 400                 405

gtc ctg agt gaa cac ctg tcc ccg cca ttc cgt gag gca ctg cac gag      1480
Val Leu Ser Glu His Leu Ser Pro Pro Phe Arg Glu Ala Leu His Glu
        410                 415                 420

ctg gca cag gag atg gag ggc agc gac aag cca cag gag ctg gcc cgg      1528
Leu Ala Gln Glu Met Glu Gly Ser Asp Lys Pro Gln Glu Leu Ala Arg
    425                 430                 435

gtc tgc ttg ggc cag gcc aat cgc cac ttt ggc atg gcg ctt ggc gcc      1576
Val Cys Leu Gly Gln Ala Asn Arg His Phe Gly Met Ala Leu Gly Ala
440                 445                 450                 455

ctc ttt gta cat gag cac ttc tca gct gcc agc aaa gcc aag gtg cag      1624
Leu Phe Val His Glu His Phe Ser Ala Ala Ser Lys Ala Lys Val Gln
                460                 465                 470

cag cta gtg gaa gac atc aag tac atc ctg ggc cag cgc ctg gag gag      1672
Gln Leu Val Glu Asp Ile Lys Tyr Ile Leu Gly Gln Arg Leu Glu Glu
            475                 480                 485

ctg gac tgg atg gac gcc gag acc agg gct gct gct cgg gcc aag ctc      1720
Leu Asp Trp Met Asp Ala Glu Thr Arg Ala Ala Ala Arg Ala Lys Leu
        490                 495                 500

cag tac atg atg gtg atg gtc ggc tac ccg gac ttc ctg ctg aaa ccc      1768
Gln Tyr Met Met Val Met Val Gly Tyr Pro Asp Phe Leu Leu Lys Pro
    505                 510                 515

gat gct gtg gac aag gag tat gag ttt gag gtc cat gag aag acc tac      1816
Asp Ala Val Asp Lys Glu Tyr Glu Phe Glu Val His Glu Lys Thr Tyr
520                 525                 530                 535

ttc aag aac atc ttg aac agc atc cgc ttc agc atc cag ctc tca gtt      1864
Phe Lys Asn Ile Leu Asn Ser Ile Arg Phe Ser Ile Gln Leu Ser Val
                540                 545                 550

aag aag att cgg cag gag gtg gac aag tcc acg tgg ctg ctc ccc cca      1912
Lys Lys Ile Arg Gln Glu Val Asp Lys Ser Thr Trp Leu Leu Pro Pro
            555                 560                 565

cag gcg ctc aat gcc tac tat cta ccc aac aag aac cag atg gtg ttc      1960
Gln Ala Leu Asn Ala Tyr Tyr Leu Pro Asn Lys Asn Gln Met Val Phe
        570                 575                 580

ccc gcg ggc atc ctg cag ccc acc ctg tac gac cct gac ttc cca cag      2008
Pro Ala Gly Ile Leu Gln Pro Thr Leu Tyr Asp Pro Asp Phe Pro Gln
    585                 590                 595
```

```
tct ctc aac tac ggg ggc atc ggc acc atc att gga cat gag ctg acc      2056
Ser Leu Asn Tyr Gly Gly Ile Gly Thr Ile Ile Gly His Glu Leu Thr
600                 605                 610                 615

cac ggc tac gac gac tgg ggg ggc cag tat gac cgc tca ggg aac ctg      2104
His Gly Tyr Asp Asp Trp Gly Gly Gln Tyr Asp Arg Ser Gly Asn Leu
                620                 625                 630

ctg cac tgg tgg acg gag gcc tcc tac agc cgc ttc ctg cga aag gct      2152
Leu His Trp Trp Thr Glu Ala Ser Tyr Ser Arg Phe Leu Arg Lys Ala
            635                 640                 645

gag tgc atc gtc cgt ctc tat gac aac ttc act gtc tac aac cag cgg      2200
Glu Cys Ile Val Arg Leu Tyr Asp Asn Phe Thr Val Tyr Asn Gln Arg
        650                 655                 660

gtg aac ggg aaa cac acg ctt ggg gag aac atc gca gat atg ggc ggc      2248
Val Asn Gly Lys His Thr Leu Gly Glu Asn Ile Ala Asp Met Gly Gly
    665                 670                 675

ctc aag ctg gcc tac cac gcc tat cag aag tgg gtg cgg gag cac ggc      2296
Leu Lys Leu Ala Tyr His Ala Tyr Gln Lys Trp Val Arg Glu His Gly
680                 685                 690                 695

cca gag cac cca ctt ccc cgg ctc aag tac aca cat gac cag ctc ttc      2344
Pro Glu His Pro Leu Pro Arg Leu Lys Tyr Thr His Asp Gln Leu Phe
                700                 705                 710

ttc att gcc ttt gcc cag aac tgg tgc atc aag cgg cgg tcg cag tcc_     2392
Phe Ile Ala Phe Ala Gln Asn Trp Cys Ile Lys Arg Arg Ser Gln Ser_
            715                 720                 725

atc tac ctg cag gtg ctg act gac aag cat gcc cct gag cac tac agg      2440
Ile Tyr Leu Gln Val Leu Thr Asp Lys His Ala Pro Glu His Tyr Arg
        730                 735                 740

gtg ctg ggc agt gtg tcc cag ttt gag gag ttt ggc cgg gct ttc cac      2488
Val Leu Gly Ser Val Ser Gln Phe Glu Glu Phe Gly Arg Ala Phe His
    745                 750                 755

tgt ccc aag gac tca ccc atg aac cct gcc cac aag tgt tcc gtg tgg      2536
Cys Pro Lys Asp Ser Pro Met Asn Pro Ala His Lys Cys Ser Val Trp
760                 765                 770                 775

tgagcctggc tgcccgcctg cacgccccca ctgcccccgc acgaatcacc tcctgctggc   2596
taccggggca ggcatgcacc cggtgccagc cccgctctgg gcaccacctg ccttccagcc   2656
cctccaggac ccggtccccc tgctgccccc cacttcagga ggggcctgga gcagggtgag   2716
gctggacttt ggggggctgt gagggaaata tactggggtc cccagattct gctctaaggg   2776
ggccagaccc tctgccaggc tggattgtac gggccccacc ttcgctgtgt tcttgctgca   2836
aagtctggtc aataaatcac tgcactgtta aaaaaaaaaa aaaaaaaaaa ttcctgcg     2894
```

<210> SEQ ID NO:2
<211> LENGTH: 775
<212> TYPE: PRT
<213> ORGANISM:Homo sapien

<400> SEQ ID NO:2

```
Met Glu Pro Pro Tyr Ser Leu Thr Ala His Tyr Asp Glu Phe Gln Glu
 1               5                  10                  15
Val Lys Tyr Val Ser Arg Cys Gly Ala Gly Gly Ala Arg Gly Ala Ser
            20                  25                  30
```

```
Leu Pro Pro Gly Phe Pro Leu Gly Ala Ala Arg Ser Ala Thr Gly Ala
          35                  40                  45
Arg Ser Gly Leu Pro Arg Trp Asn Arg Arg Glu Val Cys Leu Leu Ser
      50                  55                  60
Gly Leu Val Phe Ala Ala Gly Leu Cys Ala Ile Leu Ala Ala Met Leu
65                  70                  75                  80
Ala Leu Lys Tyr Leu Gly Pro Val Ala Ala Gly Gly Gly Ala Cys Pro
                85                  90                  95
Glu Gly Cys Pro Glu Arg Lys Ala Phe Ala Arg Ala Ala Arg Phe Leu
            100                 105                 110
Ala Ala Asn Leu Asp Ala Ser Ile Asp Pro Cys Gln Asp Phe Tyr Ser
        115                 120                 125
Phe Ala Cys Gly Gly Trp Leu Arg Arg His Ala Ile Pro Asp Asp Lys
    130                 135                 140
Leu Thr Tyr Gly Thr Ile Ala Ala Ile Gly Glu Gln Asn Glu Glu Arg
145                 150                 155                 160
Leu Arg Arg Leu Leu Ala Arg Pro Gly Gly Gly Pro Gly Gly Ala Ala
                165                 170                 175
Gln Arg Lys Val Arg Ala Phe Phe Arg Ser Cys Leu Asp Met Arg Glu
            180                 185                 190
Ile Glu Arg Leu Gly Pro Arg Pro Met Leu Glu Val Ile Glu Asp Cys
        195                 200                 205
Gly Gly Trp Asp Leu Gly Gly Ala Glu Glu Arg Pro Gly Val Ala Ala
    210                 215                 220
Arg Trp Asp Leu Asn Arg Leu Leu Tyr Lys Ala Gln Gly Val Tyr Ser
225                 230                 235                 240
Ala Ala Ala Leu Phe Ser Leu Thr Val Ser Leu Asp Asp Arg Asn Ser
                245                 250                 255
Ser Arg Tyr Val Ile Arg Ile Asp Gln Asp Gly Leu Thr Leu Pro Glu
            260                 265                 270
Arg Thr Leu Tyr Leu Ala Gln Asp Glu Asp Ser Glu Lys Ile Leu Ala
        275                 280                 285
Ala Tyr Arg Val Phe Met Glu Arg Val Leu Ser Leu Leu Gly Ala Asp
    290                 295                 300
Ala Val Glu Gln Lys Ala Gln Glu Ile Leu Gln Val Glu Gln Gln Leu
305                 310                 315                 320
Ala Asn Ile Thr Val Ser Glu Tyr Asp Asp Leu Arg Arg Asp Val Ser
                325                 330                 335
Ser Met Tyr Asn Lys Val Thr Leu Gly Gln Leu Gln Lys Ile Thr Pro
            340                 345                 350
His Leu Arg Trp Lys Trp Leu Leu Asp Gln Ile Phe Gln Glu Asp Phe
        355                 360                 365
Ser Glu Glu Glu Glu Val Val Leu Leu Ala Thr Asp Tyr Met Gln Gln
    370                 375                 380
Val Ser Gln Leu Ile Arg Ser Thr Pro His Arg Val Leu His Asn Tyr
385                 390                 395                 400
Leu Val Trp Arg Val Val Val Val Leu Ser Glu His Leu Ser Pro Pro
                405                 410                 415
Phe Arg Glu Ala Leu His Glu Leu Ala Gln Glu Met Glu Gly Ser Asp
            420                 425                 430
Lys Pro Gln Glu Leu Ala Arg Val Cys Leu Gly Gln Ala Asn Arg His
        435                 440                 445
Phe Gly Met Ala Leu Gly Ala Leu Phe Val His Glu His Phe Ser Ala
    450                 455                 460
Ala Ser Lys Ala Lys Val Gln Gln Leu Val Glu Asp Ile Lys Tyr Ile
465                 470                 475                 480
Leu Gly Gln Arg Leu Glu Glu Leu Asp Trp Met Asp Ala Glu Thr Arg
                485                 490                 495
Ala Ala Ala Arg Ala Lys Leu Gln Tyr Met Met Val Met Val Gly Tyr
            500                 505                 510
```

```
Pro Asp Phe Leu Leu Lys Pro Asp Ala Val Asp Lys Glu Tyr Glu Phe
        515                 520                 525
Glu Val His Glu Lys Thr Tyr Phe Lys Asn Ile Leu Asn Ser Ile Arg
    530                 535                 540
Phe Ser Ile Gln Leu Ser Val Lys Lys Ile Arg Gln Glu Val Asp Lys
545                 550                 555                 560
Ser Thr Trp Leu Leu Pro Pro Gln Ala Leu Asn Ala Tyr Tyr Leu Pro
                565                 570                 575
Asn Lys Asn Gln Met Val Phe Pro Ala Gly Ile Leu Gln Pro Thr Leu
            580                 585                 590
Tyr Asp Pro Asp Phe Pro Gln Ser Leu Asn Tyr Gly Gly Ile Gly Thr
        595                 600                 605
Ile Ile Gly His Glu Leu Thr His Gly Tyr Asp Asp Trp Gly Gly Gln
    610                 615                 620
Tyr Asp Arg Ser Gly Asn Leu Leu His Trp Trp Thr Glu Ala Ser Tyr
625                 630                 635                 640
Ser Arg Phe Leu Arg Lys Ala Glu Cys Ile Val Arg Leu Tyr Asp Asn
                645                 650                 655
Phe Thr Val Tyr Asn Gln Arg Val Asn Gly Lys His Thr Leu Gly Glu
            660                 665                 670
Asn Ile Ala Asp Met Gly Gly Leu Lys Leu Ala Tyr His Ala Tyr Gln
        675                 680                 685
Lys Trp Val Arg Glu His Gly Pro Glu His Pro Leu Pro Arg Leu Lys
    690                 695                 700
Tyr Thr His Asp Gln Leu Phe Phe Ile Ala Phe Ala Gln Asn Trp Cys
705                 710                 715                 720
Ile Lys Arg Arg Ser Gln Ser Ile Tyr Leu Gln Val Leu Thr Asp Lys
                725                 730                 735 -
His Ala Pro Glu His Tyr Arg Val Leu Gly Ser Val Ser Gln Phe Glu
            740                 745                 750
Glu Phe Gly Arg Ala Phe His Cys Pro Lys Asp Ser Pro Met Asn Pro
        755                 760                 765
Ala His Lys Cys Ser Val Trp
    770                 775
```

<210> SEQ ID NO:3
<211> LENGTH: 495
<212> TYPE: DNA
<213> ORGANISM:Homo sapien

<220> FEATURE:
<223> OTHER INFORMATION: n=T,C,A or G (any)

<400> SEQ ID NO:3

```
ctttgcagca agaacacagc gaaggtgggg cccgtacaat ccagcctggc agagggtctg      60
gccccttag agcagaatct ggggacccca gtatatttcc ctcacagccc cccaaagtcc      120
agcctcaccc tgctccaggc ccctcctgaa gtgaggggca gcaggggac cgggtcctgg      180
aggggctgga aggcaggtgg tgcccagagc ggggctggca ccgggtgcat gcctgccccg      240
gtagccagca ggaggtgatt cgtgcggggg cagtgggggc gtgcaggcgg gcagcaaggc      300
tcaccacacg gaacacttgt gggcagggtt catgggtgag tccttgggac agtggaaagc      360
ccggccaaac tcctcaaact gngacacact gcccagcacc ctgtagtgct caggggcatg      420
cttgtcagtc agcaactgca ggtagatgga ctgcgaccgc gcttgatgca cagttttggg      480
aaaaggcatt aaaag                                                       495
```

<210> SEQ ID NO:4
<211> LENGTH: 606
<212> TYPE: DNA
<213> ORGANISM:Homo sapien

<220> FEATURE:
<223> OTHER INFORMATION: n=T,C,A or G (any)

<400> SEQ ID NO:4
ggtcggctac ccggacttcc tgctgaaacc cgatgctgtg gacaaggagt atgagtttga 60
ggtccatgag aagacctact tcaagaacat cttgaacagc atccgcttca gcatccagct 120
ctcagttaag aagattcggc aggaggtgga caagtccacg tggctgctcc ccccacaggc 180
gctcaatgcc tactatctac ccaacaagaa ccagatggtg ttccccgcgg gcatcctgca 240
gcccaccctg tacgaccctg acttcccaca gtctctcaac tacgggggca tcggcaccat 300
cattggacat gagctgaccc acggctacga cggactgggg gggccagtat gaccgctcag 360
ggaacctgct tgcactggtg ggacggaggc ttccttacag ccgntttcct gcgaaaggct 420
gagtgcatcg ttccctnttt tatggacaac tttcaatgtn ttacaaccag gcggtgaacg 480
ggaaacacan gtttgggaga acatcgcagt atggggcggn cttaagttgg cttaccacgc 540
tattagagtt ggttncggga nggccccagg agcaccattt cccggttaaa tacanactga 600
accagt 606

<210> SEQ ID NO:5
<211> LENGTH: 352
<212> TYPE: DNA
<213> ORGANISM:Homo sapien

<220> FEATURE:
<223> OTHER INFORMATION: n=T,C,A or G (any)

<400> SEQ ID NO:5
ttttttttttt taacagtgca gtgatttatt gaccagactt tgcagcaaga acacagcgaa 60
ggtggggccc gtacaatcca gcctggcaga gggtctggcc cccttagagc agaatctggg 120
gaccccagta tatttccctc acagcccccc aaagtccagc ctcaccctgc tccaggcccc 180
tcctgaagtg aggggcagca gggggaccgg agtcctggag gggctggaag gcaggtggtg 240
cccagagcgg ggctggcacc gggtgcatgc ctgccccggt agccagcagg aggtnattcg 300
tncgggggca gtnggggcnt gcaggcgggc ancaggnttc accacacgga ac 352

<210> SEQ ID NO:6
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:6
gaagtcaggg tcgtacaggg tg 22

<210> SEQ ID NO:7
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:7
cttgttgggt agatagtagg c 21

<210> SEQ ID NO:8
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:8
ctggatgctg aagcggatgc tgttc 25

<210> SEQ ID NO:9
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:9
cctcaaactc atactccttg tccac 25

<210> SEQ ID NO:10
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:10
ccttgttgta catggagctg acatc 25

<210> SEQ ID NO:11
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:11
ccttctgttc cacagcgtct gcac 24

<210> SEQ ID NO:12
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:12
ggtgagccca tcctggtcaa tgcg 24

<210> SEQ ID NO:13
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:13
gctcggctgc gctgcggctc ag 22

<210> SEQ ID NO:14
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:14
ggtcctggag gggctggaag g 21

<210> SEQ ID NO:15
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:15
catcccgtag cccaggtggc 20

<210> SEQ ID NO:16
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:16
gccagcagga ggtgattcgt gcg 23

<210> SEQ ID NO:17
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:17
gtccccctgc tgcccctcac ttcag 25

<210> SEQ ID NO:18
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:18
ccagactttg cagcaagaac acagc 25

<210> SEQ ID NO:19
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:19
cttcctgctg aaacccgatg c 21

<210> SEQ ID NO:20
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM:Artificial Sequence

<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQ ID NO:20
ccagactttg cagcaagaac acagc 25

<210> SEQ ID NO:21
<211> LENGTH: 1092
<212> TYPE: DNA
<213> ORGANISM:Homo sapien

<400> SEQ ID NO:21
cttcctgctg aaacccgatg ctgtggacaa ggagtatgag tttgaggtcc atgagaagac 60
ctacttcaag aacatcttga acagcatccg cttcagcatc cagctctcag ttaagaagat 120
tcggcaggag gtggacaagt ccacgtggct gctcccccca caggcgctca atgcctacta 180
tctacccaac aagaaccaga tggtgttccc cgcgggcatc ctgcagccca ccctgtacga 240
ccctgacttc ccacagtctc tcaactacgg gggcatcggc accatcattg gacatgagct 300
gacccacggc tacgacgact gggggggcca gtatgaccgc tcagggaacc tgctgcactg 360
gtggacggag gcctcctaca gccgcttcct gcgaaaggct gagtgcatcg tccgtctcta 420
tgacaacttc actgtctaca accagcgggt gaacgggaaa cacacgcttg gggagaacat 480
cgcagatatg ggcggcctca agctggccta ccacgcctat cagaagtggg tgcgggagca 540
cggcccagag cacccacttc cccggctcaa gtacacacat gaccagctct tcttcattgc 600
ctttgcccag aactggtgca tcaagcggcg gtcgcagtcc atctacctgc aggtgctgac 660
tgacaagcat gcccctgagc actacagggt gctgggcagt gtgtcccagt ttgaggagtt 720
tggccgggct ttccactgtc ccaaggactc acccatgaac cctgcccaca agtgttccgt 780
gtggtgagcc tggctgcccg cctgcacgcc cccactgccc ccgcacgaat cacctcctgc 840
tggctaccgg ggcaggcatg cacccggtgc cagccccgct ctgggcacca cctgccttcc 900
agcccctcca ggacccggtc cccctgctgc ccctcacttc aggaggggcc tggagcaggg 960
tgaggctgga ctttgggggg ctgtgaggga aatatactgg ggtccccaga ttctgctcta 1020
agggggccag accctctgcc aggctggatt gtacgggccc caccttcgct gtgttcttgc 1080
tgcaaagtct gg 1092

<210> SEQ ID NO:22
<211> LENGTH: 2894
<212> TYPE: DNA
<213> ORGANISM:Homo sapien

<400> SEQ ID NO:22
ccgccgcccg cgaccctctg tggcctgcgg gcgagccgac gcgacgccga gtccgggggc 60
grgcccgggc tgggcgagcc agtggcggcc gagcccgcgc gtggacggcc gacgccgggg 120
tcccggtacg cctccgggtg ctcctccggc cgccggtgcg cgtagggcat cgggtccacc 180
gggtccagac gtggcgccgc cggagccgcg gtacctcggg ggcataagcg actgccgcgt 240
gatgctactc aaggttctcc agttcatgca ctcggcgacg ccgcgccccc cgcgcgcgcc 300
ccggagggac gggggcccga agggcaaccc gcgacgggcg tcgcggtggc cccgggccag 360
gcccgacggc gcgaccttgg ccgcgctcca cacggacgac agccccgacc acaagcggcg 420
gccggagacg cggtaagacc gccgatacga ccgggagttc atggacccgg gccagcgccg 480
gccgccgccg cggacagggc tcccgacggg actcgcgttc cggaagcgcg cgcggcgagc 540
gaaggaccgg cggttggacc tgcggtcgta gctgggtacg gtcctgaaga tgagcaagcg 600
gacgccgcca accgacgccg cggtgcggta ggggctgctg ttcgagtgga taccgtggta 660
gcgccggtag ccgctcgttt tgctcctcgc ggatgccgcg gacgaccgcg ccgggccccc 720
acccggaccg ccgcgccggg tcgcgttcca cgcgcggaag aaggcgagca cggagctgta 780

```
cgcgctctag ctcgctgacc cgggcgctgg gtacgatctc cagtagctcc tgacgccccc      840
gaccctggac ccgccgcgcc tcctcgcagg cccccagcgc cgcgctaccc tggagttggc      900
cgacgacatg ttccgcgtcc cgcacatgtc gcggcggcgc gagaagagcg agtgccagtc      960
ggacctgctg tccttgagga gcgcgatgca gtaggcgtaa ctggtcctac ccgagtggga     1020
cggtctctcc tgggacatgg agcgagtcct actcctgtca ctcttctagg accgtcgtat     1080
gtcccacaag tacctcgctc acgagtcgga ggacccacgt ctgcgacacc ttgtcttccg     1140
ggttctctag gacgttcacc tcgtcgtcga ccggttgtag tgacacagtc tcatactgct     1200
ggatgccgct ctacagtcga ggtacatgtt gttccactgc gaccccgtcg acgtcttcta     1260
gtgggggtg aacgccacct tcaccgacga tctggtctag aaggtcctcc tgaagagtct     1320
ccttctcctc caccacgacg accgctgtct gatgtacgtc gtccacagcg tcgagtaggc     1380
gaggtgtggg gtggcccagg acgtgttgat ggaccacacc gcgcaccacc accaggactc     1440
acttgtggac aggggcggta aggcactccg tgacgtgctc gaccgtgtcc tctacctccc     1500
gtcgctgttc ggtgtcctcg accgggccca gacgaacccg gtccggttag cggtgaaacc     1560
gtaccgcgaa ccgcgggaga aacatgtact cgtgaagagt cgacggtcgt ttcggttcca     1620
cgtcgtcgat caccttctgt agttcatgta ggacccggtc gcggacctcc tcgacctgac     1680
ctacctgcgg ctctggtccc gacgacgagc ccggttcgag gtcatgtact accactacca     1740
gccgatgggc ctgaaggacg actttgggct acgacacctg ttcctcatac tcaaactcca     1800
ggtactcttc tggatgaagt tcttgtagaa cttgtcgtag gcgaagtcgt aggtcgagag     1860
tcaattcttc taagccgtcc tccacctgtt caggtgcacc gacgaggggg gtgtccgcga     1920
gttacggatg atagatgggt tgttcttggt ctaccacaag gggcgcccgt aggacgtcgg     1980
gtgggacatg ctgggactga agggtgtcag agagttgatg cccccgtagc cgtggtagta     2040
acctgtactc gactgggtgc cgatgctgct gacccccccg gtcatactgg cgagtccctt     2100
ggacgacgtg accacctgcc tccggaggat gtcggcgaag gacgctttcc gactcacgta     2160
gcaggcagag atactgttga agtgacagat gttggtcgcc cacttgccct ttgtgtgcga     2220
acccctcttg tagcgtctat acccgccgga gttcgaccgg atggtgcgga tagtcttcac     2280
ccacgccctc gtgccgggtc tcgtgggtga aggggccgag ttcatgtgtg tactggtcga     2340
gaagaagtaa cggaaacggg tcttgaccac gtagttcgcc gccagcgtca ggtagatgga     2400
cgtccacgac tgactgttcg tacggggact cgtgatgtcc cacgacccgt cacacagggt     2460
caaactcctc aaaccggccc gaaaggtgac agggttcctg agtgggtact tgggacgggt     2520
gttcacaagg cacaccactc ggaccgacgg gcggacgtgc gggggtgacg ggggcgtgct     2580
tagtggagga cgaccgatgg ccccgtccgt acgtgggcca cggtcggggc gagacccgtg     2640
gtggacggaa ggtcggggag gtcctgggcc agggggacga cggggagtga agtcctcccc     2700
ggacctcgtc ccactccgac ctgaaacccc ccgacactcc ctttatatga ccccaggggt     2760
ctaagacgag attcccccgg tctgggagac ggtccgacct aacatgcccg gggtggaagc     2820
gacacaagaa cgacgtttca gaccagttat ttagtgacgt gacaattttt ttttttttt      2880
ttttttaagg acgc                                                       2894
```